GB1025 S383S2 1976

DRAFT

SANTA CLARA VALLEY WATER DISTRICT

DEVELOPMENT OF REGIONAL REGRESSION EQUATIONS FOR SOLUTION OF CERTAIN HYDROLOGIC PROBLEMS IN AND ADJACENT TO SANTA CLARA COUNTY

DRAFT

by

A. D. Saah, P.E.
Associate Civil Engineer (Hydrologist)

and

Randy Talley, P.E.
Assistant Civil Engineer

W. James Sanchez, Jr., P.E.
Assistant Civil Engineer

under the direction of
Eugene Watson
Division Engineer
Resources Management Division

November 1976

DRAFT

DAVID K. GILL
Advanced Planning Manager

LLOYD C. FOWLER
Chief Engineer

DISTRICT BOARD OF DIRECTORS

Joe Chiri	District 1	James J. Lenihan, Chairman	District 5
Patrick T. Ferraro	District 2	Courtland M. Rush, Vice Chairman	At Large
Robert T. Sapp	District 3	Linda Peralta	At Large
Maurice E. Dullea	District 4		

DRAFT

TABLE OF CONTENTS

Page

Introduction 1

Results 3

Physical Setting 5

Stochastic Generation and Statistical Analysis of Data 8

- Basic Data 8
- Data Analysis 11
  - Grouping 11
  - Analysis of Computer Output 13
- Regional Regression Analysis 19
  - Independent Variables 19
  - Regression and Correlation 21
  - Review of the Procedures 27
- Preliminary Selection of Stations and Their Respective Significant Physical Parameters 29
  - Regression for the Mean, M.................. 29
  - Regression for the Standard Deviation 32
  - Regression for the 1% Peak Directly 33
- Determination of Final Regression Equations 36
  - Small Basins 43

Determination of Skew Coefficient 44

Confidence Limits 48

Application of Regional Equations 52

- General 52
- Equivalent Years of Record 52
- Sensitivity Analysis 53
- Volumes Equations 55
- Range of Application 56

References 57

Appendix A* - Basic Data

Appendix B* = Statistical Analysis (Computer Printouts).

Appendix C* - Regression Analysis (Computer Printouts) .

Appendix D - Volumes Equations 58

Appendix E- Frequency Curves 71

---

*Available upon request

INTRODUCTION

The work reported on here was undertaken to improve, update and fully document the Santa Clara Valley Water District's regional regression equations. The regional equations are an important part of the District's methodology for determining design flood flows and their frequency of occurrence in Santa Clara County.

Over the past ten years the District has utilized regional regression and correlation techniques to estimate design flood flows and their frequency of occurrence. In its endeavors to provide the best available methods, the District updates these regional regression equations periodically as more data becomes available (usually every five years). The year 1976 was the year during which this updating was done. Timely with that, the District convened a panel of expert hydrologists to review its methods. The panel reviewed, in their report (Review of Basic Hydrology Methods for Flood Control), the current methods and recommended short- and long-term improvements. This study also incorporates the panel's recommendations.

The report discusses:

1. The available basic streamflow data and associated problems.

2. The stochastic generation of data and statistical frequency analysis.

3. The determination of the physical and hydrometeorological characteristics of the watersheds.

4. The multiple liner regression analysis used to determine the regional ~~regional~~ regression equations, and

5. The application and limitations of the developed equations.

RESULTS

This study has improved the predictive capability of the regional regression equations. With the inclusion of new stations and up-to-date records, the new equations become more reliable and applicable over a wider range of conditions. There are separate regression equations for mean and σ which together define flow frequency regime.

The mean can be estimated fairly well from a limited amount of data so no large changes were expected nor were found in this investigation.

The greatest progress made in this study was in the development of the regression equation for the standard deviation. The coefficient of determination improved substantially with a very small standard error (.065).

The regression constant was reduced and greater weight was put on the Mean Annual Precipitation and less on the area, a fact which made these equations more sensitive to the variability in the physical and climatological conditions.

In this study, equivalent years of record at all utilized stations were based on a 61-year regionally-correlated data matrix. The use of the 20-station, 61-year data matrix is responsible for the improvement in defining the flow variability. The correlation and extension of each station's data provided a more uniform base upon which to perform regressions.

Although the regional equations are most useful on ungaged natural watersheds, they may also be used to augment information on watersheds where streamflow data is available. However, in situations where the regional equations are applied to gaged watersheds and don't concur, considerable additional comprehensive analysis of all available records must be undertaken. The objectives in this regard are to obtain the best possible estimate of the design flows. Thus, the results of the regional equations are to be checked and compared with data in or around the specific watershed being studied and are to be modified, if necessary, in those cases where good record exists, to more closely reflect the information inherent in the historic record.

It must be pointed out, however, that the exercise of considerable judgement by an experienced and knowledgeable hydrologist is essential to the practice of hydrology. Naive application of the methods reported on here could lead to erroneous results.

PHYSICAL SETTING

Santa Clara County is located within the west-central coast ranges of California south of San Francisco Bay. The coast range landscape is characterized by a series of northwesterly trending, sub-parallel mountain ranges and intervening large and small valleys. Santa Clara Valley, which extends 50 miles southwesterly through Santa Clara County into San Benito County, is a complex down-dropped structural trough. This alluvium-filled valley is bounded on the northeast by Diablo Range and on the southwest by Santa Cruz Mountains. The width of the valley varies from about 14 miles in the northern portion to less than one mile at the midpoint, then to about 5 miles in the southern portion of the county before widening again in San Benito County.

Santa Clara Valley is separated into two surface water drainage areas (north and south) by a subtle topographic divide in the alluvium near Morgan Hill.

Elevation of the floor of the valley ranges from sea level at San Francisco Bay to about 425 feet at the apex of the alluvial fan which forms the drainage divide, and 140 feet at the basin outlet of South Santa Clara Valley.

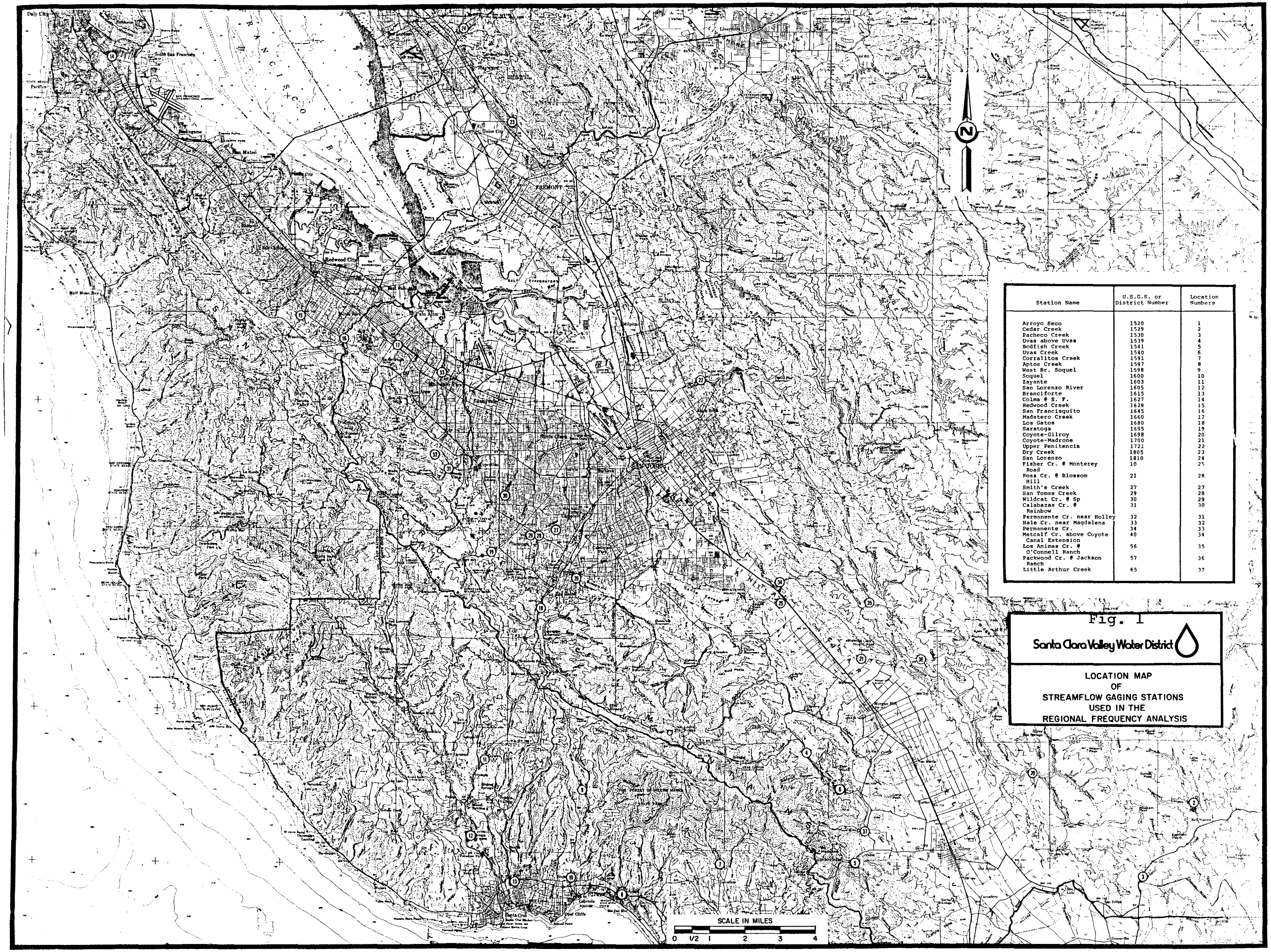
The east slopes of the Santa Cruz Mountains gradually rises from the foothills to moderately steep, rugged ridges to a peak elevation of 3798 feet at Loma Prieta.

The western slopes of the Diablo Range are neither as rugged nor as steep as the Santa Cruz Mountains but rise to a peak elevation of 4372 feet at Copernicus peak.

The mean annual precipitation in the County varies with elevation, from 14 inches in the valley floor to 56 inches in the high point of Loma Prieta. Most of the rain occurs between November and April. Storms are usually of the cyclonic type, with fronts usually covering the whole County at one time. Convective storms with heavy central rain and limited spacial distribution occur rarely. It is the activity of some cells within the cyclonic storms that cause the relatively varied distribution of rainfall in time and space.

Snow rarely occurs, but when it does it is usually very light and melts off rapidly. Occasionally significant snow falls accrue in the surrounding mountains staying for up to several weeks. However, snowmelt does not contribute significantly to flood flows.

Topographic features of the County and slope characteristics, which more strongly affect land use, and surface runoff are illustrated in Figure 1.



Station Name	U.S.G.S. or District Number	Location Numbers
Arroyo Seco	1520	1
Cedar Creek	1529	2
Pacheco Creek	1530	3
Uvas above Uvas	1539	4
Bodfish Creek	1541	5
Uvas Creek	1540	6
Corralitos Creek	1591	7
Aptos Creek	1597	8
West Br. Soquel	1598	9
Soquel	1600	10
Zayante	1603	11
San Lorenzo River	1605	12
Branciforte	1615	13
Colma @ S. F.	1627	14
Redwood Creek	1628	15
San Francisquito	1645	16
Madatero Creek	1660	17
Los Gatos	1680	18
Saratoga	1695	19
Coyote-Gilroy	1698	20
Coyote-Madrone	1700	21
Upper Penitencia	1721	22
Dry Creek	1805	23
San Lorenzo	1810	24
Fisher Cr. @ Monterey Road	10	25
Ross Cr. @ Blossom Hill	21	26
Smith's Creek	27	27
San Tomas Creek	29	28
Wildcat Cr. @ Sp	30	29
Calabazas Cr. @ Rainbow	31	30
Permanente Cr. near Holley	32	31
Hale Cr. near Magdalena	33	32
Permanente Cr.	34	33
Metcalf Cr. above Coyote Canal Extension	40	34
Los Animas Cr. @ O'Connell Ranch	56	35
Packwood Cr. @ Jackson Ranch	57	36
Little Arthur Creek	65	37

Fig. 1

Santa Clara Valley Water District

LOCATION MAP
OF
STREAMFLOW GAGING STATIONS
USED IN THE
REGIONAL FREQUENCY ANALYSIS

STOCHASTIC GENERATION AND STATISTICAL ANALYSIS OF DATA

Basic Data

Streamflow data were gathered from stations within and outside the District. During the process of extracting the peaks and the annual 1, 2, 3, 4, 5, 10, and 15-day maximum volumes, the data were screened for reliability and consistency. Figure 1 shows the location of these stations while Figure 2 shows the available record. Out of the 37 stations that were primarily considered for this study, 24 are U.S.G.S. and 13 are owned and operated by the District.

Most of the U.S.G.S. stations in Santa Clara County and some from adjacent counties were used in this study. The stations in nearby counties were selected on the basis of apparent similarities in physiographic characteristics.

Most District stations were initially considered for the analysis. Stations downstream of reservoirs or some other regulating facilities were removed from consideration, since inclusion would impose nonhomogeniety of data in the statistical computations. The remaining stations were reviewed for length of record and an arbitrary ten-year minimum was selected as a limiting criterion.

Also, only one station per watershed was used in an effort to prevent unnatural weighting toward a specific type of streamflow pattern or distribution.

The stations used for the initial analysis are listed in Table 1.

In general, the U.S.G.S. streamflow stations were more reliable than the District stations. The correlation between the District's and U.S.G.S. stations were poor at times.

Screening the obvious bad data points out of the record improved the correlation coefficients greatly as will be shown later.

For the years 1926-1960, U.S.G.S. streamflow data were obtained from "Geological Survey Water Supply Papers, Surface Water Supply of the United States". After that, 1961-1975, the data were obtained from "Water Resources Data for California, Part I. Surface Water Records". The annual peaks and the 1, 2, 3, 4, 5, 10, and 15-day volumes were extracted.

District streamflow data were obtained from the District's records. As with the U.S.G.S. stations, the annual peaks and volumes were also extracted.

A list of all the data together with the time and date of occurrence is available in Appendix A. The inconsistancies that were discovered and corrected in the basic data are also listed in the same Appendix.

TABLE 1

Station Name	U.S.G.S. or District Number	Location Numbers
Arroyo Seco	1520	1
Cedar Creek	1529	2
Pacheco Creek	1530	3
Uvas above Uvas	1539	4
Bodfish Creek	1541	5
Uvas Creek	1540	6
Corralitos Creek	1591	7
Aptos Creek	1597	8
West Br. Soquel	1598	9
Soquel	1600	10
Zayante	1603	11
San Lorenzo River	1605	12
Branciforte	1615	13
Colma @ S. F.	1627	14
Redwood Creek	1628	15
San Francisquito	1645	16
Madatero Creek	1660	17
Los Gatos	1680	18
Saratoga	1695	19
Coyote-Gilroy	1698	20
Coyote-Madrone	1700	21
Upper Penitencia	1721	22
Dry Creek	1805	23
San Lorenzo	1810	24
Fisher Cr. @ Monterey Road	10	25
Ross Cr. @ Blossom Hill	21	26
Smith's Creek	27	27
San Tomas Creek	29	28
Wildcat Cr. @ Sp	30	29
Calabazas Cr. @ Rainbow	31	30
Permanente Cr. near Holley	32	31
Hale Cr. near Magdalena	33	32
Permanente Cr.	34	33
Metcalf Cr. above Coyote Canal Extension	40	34
Los Animas Cr. @ O'Connell Ranch	56	35
Packwood Cr. @ Jackson Ranch	57	36
Little Arthur Creek	65	37

Data Analysis

This section discusses the systematic screening of data and the logical grouping of stations in order to improve correlations. The streamflow data, i.e., annual peaks and volumes for various durations, were analyzed by the use of the U.S. Army Corps of Engineers' Hydrologic Engineering Center (HEC) program titled "Regional Frequency Computation".

This program performs frequency computations of hydrologic events utilizing the information from adjacent stations. Frequency statistics are computed for recorded events at each station and for each duration. Statistics for each station are then adjusted to the complete record of selected base stations in accordance with the procedure given in "Statistical Methods in Hydrology" by Leo R. Beard, January 1962. Missing events are estimated by the use of cross-correlation matrices of groups of stations, so that complete sets of events are obtained for all years at all stations while preserving all intercorrelations. The adopted statistics of estimated and observed data at all stations were used in the multiple linear regression and correlation. A listing of the stations together with their record lengths is shown on Figure 2.

Grouping

The method of computation in the previously mentioned HEC program requires that a maximum of only ten stations be grouped together for each run. Since 37 stations were initially considered, it was not possible to cross-correlate all stations together with all durations (i.e., Peak 1, 2, 3, 4, 5, 10, and 15-day volumes). Therefore, it was necessary to "group" the stations in some logical fashion.

YEARS OF DATA AVAILABLE FOR REGIONAL FREQUENCY ANALYSIS

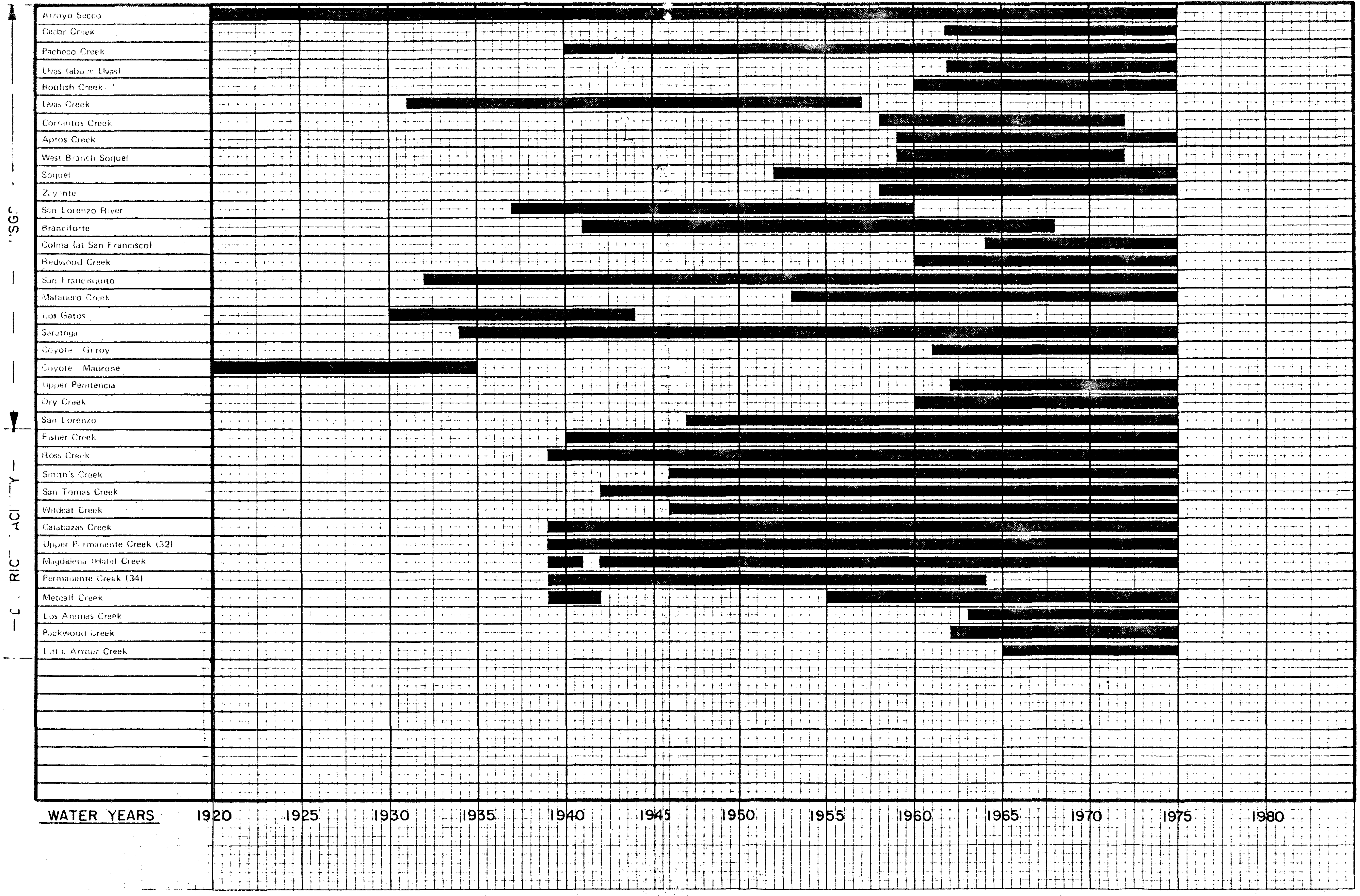


FIG. 2

The initial criteria for grouping stations were based on their geographical proximities. Preliminary grouping was done by visual inspection after first plotting the stations on a topographic map,(Figure 1).

Grouping was also necessary since Arroyo Seco near Soledad has the longest record. Hence, all the stations could be overly affected by it. The grouping process was used to alleviate the possible excessive weighing of any one station.

In this grouping a matrix of cross correlations was created. The objectives were to obtain the groups that, together, would yield the best cross correlation coefficient.

The following chapter discusses the several computer trial runs needed to obtain the best grouping.

Analysis of Computer Output

The criterion selected for determining the "goodness" of a particular station's grouping was the correlation coefficients matrices of the recorded and estimated data. Each correlation coefficient matrix and grouping for each run was carefully reviewed before any adjustments were made. Additionally, relative information ratios as well as the basic data were considered in the evaluation of a streamflow station.

The stations were arranged in runs and groups. Each run contained many groups of stations. These groups were arranged in such a way as to enable station statistics and data to be saved and used again for subsequent groups. This procedure was utilized to maintain consistency in data transfer and to tie the stations together.

The groups were arbitrarily numbered from 10 to 16. After reviewing the results of each group, subsequent runs were made with the modifications deemed necessary, and were labeled run 1, 2, 3, 4, 5, and run 6, respectively.

In the preliminary runs, only 50 years of data were used for the study. A dimension limit in the program prohibited more than 50 years of data with 8 durations (i.e., 400 data points). Extension of the necessary dimensions required a somewhat complex procedure that was not attempted. Only one station, Arroyo Seco, had an additional 20 years of record. However, for the final runs it was decided to include 11 of these 20 additional years at Arroyo Seco. Therefore, the number of durations investigated was reduced.

Visual inspection of the correlation matrix for some of the groupings showed that a few correlation coefficients were quite low. However, an analysis considering the relative information ratio as described in "Use of Correlation to Improve Estimates of the Mean and Variance" by Myron B. Fiering, U.S.G.S. Professional Paper 434-C, demonstrated that the information gained by estimating additional years of record exceeded the information loss due to the decreased correlation coefficients even at places where the correlations are relatively low. This being the case, it was considered appropriate to leave in some of the stations even though they produced low correlation coefficients.

For some unknown reason very large negative skew coefficients exist within the data. Low and high outliers were looked into according to the method described in "A Guideline for Determining

Flood Flow Frequencies", U.S. Water Resources Council. The largest flow for each station that could be classed as a low outlier was determined. (This procedure does not consider zero flows as low outliers.) Comparison of these flows with the recorded data (Table 2) revealed that there were no "low outliers" present.

Based on the correlation coefficient matrixes and considering the extra effort required to realize possible small improvements for running additional groupings, the runs were terminated and the obtained statistics (dependent variables) from run 6 were adopted and are listed on Table 3. Appendix B is the complete printout of all the basic data together with the various corrolation coefficient matrixes. The estimated data and the various steps needed to obtain the adopted statistics are clearly shown.

TABLE 2

PEAK		ANALYSIS OF OUTLIERS FOR PEAKS				
w/generalized $\bar{G}$ = -0.60		Recorded $\bar{G}$	N Years	σ	$\bar{X}$	Low Outliers Limit
Arroyo Secco	11152000	-0.575	50	0.378	3.877	205.1
Cedar Creek	11152900	-1.700	14	0.759	2.564	1.1
Pacheco Creek	11153000	-1.492	36	0.922	3.060	0.3
Uvas (above Uvas)	11153900	-1.579	14	0.303	3.384	239.2
Bodfish Creek	11154100	-0.411	16	0.479	2.455	6.7
Uvas Creek	11154000	-0.392	27	0.402	3.375	74.3
Corralitos Creek	11159150	-0.920	15	0.538	2.639	6.8
Aptos Creek	11159700	-1.125	17	0.583	2.616	4.1
West Branch Soquel	11159800	-0.820	14	0.532	2.956	15.5
Soquel	11160000	-0.930	24	0.477	3.368	41.6
Zayante	11160300	-1.253	18	0.587	2.903	7.3
San Lorenzo River	11160500	-0.344	23	0.450	3.775	137.3
Branciforte	11161500	-0.206	19	0.381	3.173	68.2
Colma (@ San Francisco)	11162720	0.936	11	0.182	3.073	4.5
Redwood Creek	11162800	-1.610	16	0.403	2.310	8.7
San Francisquito	11164500	-2.384	35	0.828	2.943	0.5
Matadero Creek	11166000	-2.518	23	0.743	2.322	0.4
Los Gatos	11168000	-0.566	15	0.441	3.261	60.0
Saratoga	11169500	0.254	42	0.379	2.621	14,040.5 *
Coyote-Gilroy	11169800	-1.676	15	0.612	3.334	18.9
Coyote-Madrone	11170000	-1.499	10	0.522	3.449	67.7
Upper Penitencia	11172100					
Dry Creek	11180500	-1.295	16	0.787	2.181	0.3
San Lorenzo	11181000	-0.472	29	0.546	2.973	8.0
Fisher Creek	10	-0.684	13	0.477	2.426	7.4
Ross Creek	21	-0.393	29	0.403	2.177	4.5
Smith's Creek	27					
San Tomas Creek	29	-2.433	26	0.691	2.089	0.3
Wildcat Creek	30	-0.837	29	0.433	2.513	7.5
Calabazas Creek	31	0.191	27	0.413	2.453	9,963.0*
Upper Permanente Creek	32	-0.308	25	0.330	1.815	4.0
Magdalena (Hale) Creek	33	-2.424	30	0.821	1.853	0.1
Permanente Creek	34	-1.790	19	0.824	2.202	0.2
Metcalf Creek	40	-0.199	18	0.505	0.720	0.1
Los Animas Creek	56	-1.646	13	0.596	2.480	3.4
Packwood Creek	57	-0.603	12	0.766	1.785	0.2
Little Arthur Creek	65					

*High Outliers

TABLE 3

REGIONAL REGRESSION ANALYSIS ADOPTED STREAM GAGE STATISTICS FOR PEAK FLOWS

		D.A.	M	σ	1%	CSM*
1.	Aptos Creek	12.2	2.621	.536	4,252	349
2.	Bodfish Creek	7.4	2.448	.488	2,319	313
3.	Branciforte Creek	17.3	3.130	.357	6,326	365
4.	Calabazas Creek SCVWD #31	4.73	2.479	.449	2,104	445
5.	Cedar Creek	12.8	2.355	.896	10,951	855
6.	Corralitos Creek	10.6	2.615	.477	3,249	306
7.	Coyote Creek near Gilroy	109.0	3.365	.560	26,169	240
8.	Coyote Creek at Madrone	196.0	3.478	.809	99,751	508
9.	Dry Creek	9.41	2.195	.766	4,315	458
10.	Fisher Creek SCVWD #10	15.5	2.263	.663	3,231	208
11.	Hale Creek SCVWD #33	2.8	1.903	.789	2,433	869
12.	Los Animas Creek SCVWD #56	37.0	2.343	.729	5,170	139
13.	Los Gatos Creek	38.6	3.268	.421	11,467	297
14.	Matadero Creek	7.26	2.456	.469	2,176	300
15.	Metcalf Creek SCVWD #40	.80	.684	.510	43	55
16.	Pacheco Creek	146.0	2.971	.917	49,540	339
17.	Packwood Creek SCVWD #57	10.1	1.652	.910	2,305	228
18.	Permanente Creek SCVWD #34	15.9	2.296	.699	4,074	256
19.	Redwood Creek	1.82	2.291	.451	1,376	756
20.	Ross Creek SCVWD #21	2.86	2.188	.413	921	322
21.	San Francisquito Creek	37.5	3.044	.551	12,019	320
22.	San Lorenzo Creek at Haywood	37.5	2.987	.579	11,899	317
23.	San Lorenzo River	111.0	3.726	.400	30,060	270
24.	San Tomas Creek SCVWD #29	3.25	2.175	.716	3,319	1021
25.	Saratoga Creek at Saratoga	9.2	2.598	.364	1,915	208

TABLE 3 (continued)

	D.A.	M	σ	1%	CSM*
26. Soquel Creek at Soquel	40.2	3.295	.411	11,686	290
27. Upper Permanente SCVWD #32	8.2	1.870	.336	317	38
28. Uvas Creek at Morgan Hill	30.3	3.391	.348	11,097	366
29. Uvas Creek above Uvas Reservoir	21.0	3.318	.364	10,053	478
30. West Branch Soquel Creek	12.2	2.974	.471	7,235	593
31. Wildcat Creek SCVWD #30	1.15	2.502	.454	2,267	1971
32. Zayante Creek	11.1	2.825	.546	7,103	639

*Cubic feet per second per square mile

Regional Regression Analysis

In the previous chapter, statistics (mean, standard deviation and 1% flows) were obtained for gaged watersheds. These statistics contain information available from all gaged watersheds as well as representing information transferable to the region. In this chapter, these statistics are used to develop equations to determine design floods and their frequency of occurrence. The mathematical model used was the Corps of Engineers, HEC Program, Multiple Linear Regression, January 1975.

In general, multiple linear regression determines the relationships between one dependent and two or more independent variables. In this case the dependent variables for the various equations are the statistics of recorded data, determined previously, while the independent variables are watershed climatological or physical parameters described below.

Independent Variables

The rate of runoff from a basin is a function of its hydrologic and physical characteristics. To describe the relation between design flows and basin characteristics, it was first necessary to determine which of the hydrologic characteristics are significantly related to the statistics of flood peaks and volumes. The choice of independent variables was limited to those readily obtainable. Thirteen of the geomorphic and climatic variables that were considered in this analysis as independent variables are described below. The actual values were obtained and are listed on Table 4.

Area, A - The area planimetered from a U.S.G.S. 7-1/2 minute quad map in square miles.

M.A.P., P - The mean annual basin precipitation determined from the District's 1972 Mean Annual Precipitation map (M.A.P.). For stations outside the limits of the District's map, the M.A.P. was estimated from the U.S.G.S. map developed for all of California.

Bifurcation Ratio, Rb - The average of the ratios of the number of stream segments of a given order to the number of stream segments of the next higher order. Stream orders were determined from designated flow lines on U.S.G.S. 7-1/2' quad maps.

Length of Main Stream, Lm - Length of longest flow path from stream gage location to furthest distant point on the drainage divide (in miles).

Shape Factor, d/Lm - The average of the distances from the points on a uniform grid (50-70 points over the basin) to the stream gage location (as measured along the flow path) divided by Lm, the length of the main stream.

Effective Basin Width, Wb - Total Area A divided by length of main stream Lm.

Slope of Main Stream, Sm - Average slope of the main stream in Ft/Ft. Profile plotted from U.S.G.S. 7-1/2 ' quad maps.

Total Stream Lengths, Ls - Total length (in miles) of all the stream paths designated on the U.S.G.S. 7-1/2 ' quad maps.

Drainage Density, DD - Total stream lengths, Ls divided by the drainage area, A.

Distance to Center of Area, Lca - Distance (in miles) from the stream gage to the center of area of the drainage basin measured along the stream path to a line drawn through the center of area perpendicular to the stream.

Overland Flow, d - An estimate of the length of the overland path (in miles) defined as the reciprocal of twice the drainage density, DD.

Basin Factor, LLS - It is determined from a combination of three measured parameters: $\frac{LmLca}{S_m^{1/2}}$.

Average Basin Slope, So - From a uniform grid of 50-70 points per basin the slope at a grid point is the average of two sides of the grid square. The average of all the slopes at each grid point is the average basin slope, So in ft/ft. Most easily calculated by using contour interval and counting contour lines.

Regression and Correlation

In this analysis it is important to recognize the differences between regression and correlation.

Regression is the mathematical method that yields equations of one dependent and one or more independent variables. Correlation is the process by which the degree of association between these variables is defined. The degree of association is a measure of how good the equations are.

Correlation theory requires that the data of the two processes to be correlated are normally distributed, although this condition should not limit the application of that theory. McDonald (1957) reported that the "non-normality is of inconsequential magnitude physically". The other condition of correlation is the assumption that the processes are free from measurement errors. Regression requires the following conditions in addition to what is required for correlation.

TABLE 4

REGIONAL REGRESSION ANALYSIS

BASIN PARAMETERS

STATION	AREA A Miles^2	M.A.P. P Inches	BIFURCATION RATIO Rb	LENGTH OF MAIN STREAM Lm Miles	SHAPE FACTOR d/Lm Ft/Ft	EFFECTIVE BASIN WIDTH Wb Miles	SLOPE OF MAIN STREAM Sm Ft/Ft	TOTAL STREAM LENGTHS Ls Miles	DRAINAGE DENSITY DD Miles/Mile^2	DISTANCE TO CENTER OF AREA Lca Miles	OVERLAND FLOW d Miles	$\frac{LmLca}{Sm^{1/2}}$ LLs	AVERAGE BASIN SLOPE So Ft/Ft
1159700 Aptos Creek	12.2	36	4.00	9.09	.550	1.34	.0235	21.93	1.80	4.74	.278	280.8	.2390
1154100 Bodfish Creek	7.4	30	3.33	5.68	.455	1.30	.0327	14.58	1.97	1.89	.254	59.4	.2412
1161500 Branciforte Creek	17.3	40	3.50	9.09	.558	1.90	.0172	31.63	1.83	4.74	.274	328.3	----
SCVWD #31 Calabazas Creek	4.73	24	3.2	5.11			.0278			2.73		83.6	
1152900 Cedar Creek near Bell Station	12.8	21	2.83	8.14	.518	1.57	.0271	27.27	2.13	3.98	.235	196.8	.2376
1159150 Corralitos Creek	10.6	40	4.25	5.76	.602	1.84	.0395	20.8	1.97	3.03	.254	87.8	----
1169800 Coyote Creek near Gilroy	109.0	24	3.68	28.41	.440	3.84	.0141	182.58	1.68	10.04	.299	2405.0	----
1170000 Coyote Creek at Madrone	196.0	22	4.17	41.29	.451	4.75	.0098	343.56	1.75	10.98	.285	4575.8	----

TABLE 4 (CONTINUED)

REGIONAL REGRESSION ANALYSIS

BASIN PARAMETERS

STATION	AREA A Miles^2	M.A.P. P Inches	BIFURCATION RATIO Rb	LENGTH OF MAIN STREAM Lm Miles	SHAPE FACTOR d/Lm Ft/Ft	EFFECTIVE BASIN WIDTH Wb Miles	SLOPE OF MAIN STREAM Sm Ft/Ft	TOTAL STREAM LENGTHS Ls Miles	DRAINAGE DENSITY DD Miles/Mile^2	DISTANCE TO CENTER OF AREA Lca Miles	OVERLAND FLOW d Miles	$\frac{LmLca}{Sm^{1/2}}$ LLs	AVERAGE BASIN SLOPE So Ft/Ft
1180500 Dry Creek @ Union City	9.41	19											
SCVWD #10 Fisher Creek	15.5	20	2.23	8.14	-	1.90	.0025	23.5	1.52	3.79	.330	617.0	-
SCVWD #33 Hale Creek	2.8	19	4	2.12	.565	1.32	.0308	5.11	1.83	1.00	.273	12.1	.1040
SCVWD #56 Los Animas Creek	37.0	20	3.80	14.39	.546	2.57	.0146	89.20	2.41	6.06	.207	721.6	-
1168000 Los Gatos Creek	38.6	43	4.11	12.88	.410	3.00	.0260	82.58	2.14	6.44	.234	514.0	.2810
1166000 Matadero Creek	7.26	19	7	5.91	.573	1.23	.0138	13.03	1.79	3.22	.279	162.0	.1008
SCVWD #40 Metcalf Creek	.80	17	2	2.05	-	.390	.1074	2.65	3.31	1.04	.151	6.5	-
1153000 Pacheco Creek	146.0	21											
SCVWD #57 Packwood Creek	10.1	20	4.67	7.01	.633	1.44	.0319	22.35	2.21	4.36	.226	171.2	-

TABLE 4 (CONTINUED)

REGIONAL REGRESSION ANALYSIS

BASIN PARAMETERS

STATION	AREA A Miles^2	M.A.P. P Inches	BIFURCATION RATIO Rb	LENGTH OF MAIN STREAM Lm Miles	SHAPE FACTOR d/Lm Ft/Ft	EFFECTIVE BASIN WIDTH Wb Miles	SLOPE OF MAIN STREAM Sm Ft/Ft	TOTAL STREAM LENGTHS Ls Miles	DRAINAGE DENSITY DD Miles/Mile^2	DISTANCE TO CENTER OF AREA Lca Miles	OVERLAND FLOW d Miles	$\frac{LmLca}{Sm^{1/2}}$ LLs	AVERAGE BASIN SLOPE So Ft/Ft
1172100 Penitencia Creek	21.1	21.5	3.06	11.93	.555	1.77	.0311	41.29	1.96	10.61	.256	717.6	.2479
SCVWD #34 Permanente	15.9	21	4.13	9.85	.483	1.61	.0226	28.22	1.77	4.73	.282	309.9	----
11628000 Redwood Creek	1.82	9	2.00	2.08	.630	0.87	.0236	2.46	1.35	1.14	.370	15.4	.1233
SCVWD #21 Ross Creek	2.86	25	3	1.89	.600	1.51	.0360	3.79	1.33	.83	.376	8.3	.1669
1164500 San Francisquito Creek	37.5	31	3.71	11.74	.578	3.19	.0129	73.86	1.97	5.87	.254	606.7	----
1181000 San Lorenzo Creek at Hayward	37.5	24	4.10	8.79	.6353	4.27	.0152	80.30	2.14	4.55	.234	324.1	.2168
1160500 San Lorenzo River	111.0	48	3.66	21.21	.5379	5.23	.0071	135.04	1.22	10.04	.411	2519.6	----
SCVWD #29 San Tomas Creek	3.64	29	2.00	5.42			.0385			2.73			

TABLE 4 (CONTINUED)

REGIONAL REGRESSION ANALYSIS

BASIN PARAMETERS

STATION	AREA A Miles^2	M.A.P. P Inches	BIFURCATION RATIO Rb	LENGTH OF MAIN STREAM Lm Miles	SHAPE FACTOR d/Lm Ft/Ft	EFFECTIVE BASIN WIDTH Wb Miles	SLOPE OF MAIN STREAM Sm Ft/Ft	TOTAL STREAM LENGTHS Ls Miles	DRAINAGE DENSITY DD Miles/Mile^2	DISTANCE TO CENTER OF AREA Lca Miles	OVERLAND FLOW d Miles	$\frac{LmLca}{Sm^{1/2}}$ LLs	AVERAGE BASIN SLOPE So Ft/Ft
1169500 Saratoga Creek at Saratoga	9.2	42.0	2.85	5.11	.610	1.80	.0500	9.47	1.03	2.84	.485	64.9	.2765
1160000 Soquel Creek at Soquel	40.2	40	3.31	15.91	.437	2.53	.015	62.92	1.57	8.90	.320	1156.2	----
SCVWD #32 Upper Permanente	8.2	24	3.67	7.01	.515	1.17	.0354	18.37	2.24	3.60	.223	134.0	.2376
1154000 Uvas Creek at Morgan Hill	30.3	37.4	4.02	10.98	.515	2.76	.0207	59.09	1.95	5.34	.256	407.6	.2448
1153900 Uvas Creek above Uvas Reservoir	21.0	39.0	3.83	8.90	.767	2.36	.0234	43.18	2.06	3.79	.243	220.5	.2353
1159800 West Branch Soquel Creek	12.2	45	2.50	9.28	.499	1.32	.0224	16.89	1.82	3.98	.275	246.8	.244
SCVWD #30 Wildcat Creek	4.12	2[illegible]	2.50	4.73			.0320			2.58		68.2	
1160300 Zayante Creek	11.1	50	4.00	6.82	.6165	1.63	.0236	19.89	1.79	3.41	.279	151.4	----

1. The deviations of the dependent variable about the regression line are normally distributed.

2. Values of dependent variables are uncorrelated.

3. The variables are homogenious, i.e., variables should be measured the same way, and statistically independent. The assumption of normality is required for the direct application of the standard error theories in statistics and the computation of confidence limits and tests of significance.

It is evident that the hydrologic data used in regression and correlation cannot, by any means, satisfy all of these assumptions. Therefore, the resultant regression equations should be used under the implied assumption, that personal experience in interpreting the results should be utilized as a major input.

For these reasons, and in order to limit and guide the personal judgement a confidence band (as will be discussed later) should always be associated with the results of regression operations.

The log transformation of the dependent variables (Table 3) and the independent variables (Table 4) was used in the regression. Although this transformation could distort basic information, it was thought that the benefits of the log transformation as a process of normalizing the data and linearizing the regression model to provide equal deviation about the regression line ,far exceed the small loss of information.

Stepwise multiple linear regression is a satisfactory technique for eliminating variables of less significance, if the independent variables are carefully selected on a logical basis. The use of too many variables reduces the degrees of freedom, and, hence, the correlation coefficient. Difficult to measure and relatively unimportant independent variables together with variables of less significance were not considered in the final results of the regression.

The following is the form of the general regression equation:

$\text{Log } Y = a \log A + b \log B + c + --- + k$

Where Y is the dependent variables

A,B,C--- are independent variables,

a,b,c--- are regression coefficients and

k is the regression constant.

The regression equation of the above form is derived by least squares fitting, the result of which is an establishment of values of regression constants and coefficients in such a way that the determination coefficients are maximized and the standard errors minimized.

Elimination of independent variables by the stepwise method was done based on the smallest partial determination coefficient.

Review of the Procedures

Among other statistics, the results of the HEC regression program consists of the following information (reference is made to Appendix C):

1. Simple correlation coefficients between parts of independent variables. One of the practical requisites in multiple-regression analysis is that the various independent variables not be highly correlated amongst themselves. It is therefore, the intention of this study to choose the variables in such a way that this measure is minimized.

2. Partial determination coefficient. It is the measure used to eliminate the least significant independent variable in the stepwise process.

3. Regression coefficient for each independent variable.

4. Regression Constant.

5. Multiple determination coefficient both actual and adjusted. The adjustment is used to remove the bias introduced by less degress of freedom. The maximization of this function is one measure used in the final choice of the independent variables.

6. The standard error is the last measure, the minimization of which is used in the choice of the independent variables.

The HEC regression program eliminates the least significant independent variable in the computed regression equation, then proceeds to recompute successive equations without that variable. This procedure was repeated until only one independent variable remains. This made it possible to determine when no significant improvement in the results was obtained by the use of additional variables.

Normally, the independent variables, drainage area and precipitation, explain most of the explainable variation in flood magnitude. Other variables should be used only when they improve results by an amount that justifies the necessary data collection.

Of the 13 original variables, three were eliminated immediately because the numerical determination of them was difficult and preliminary regression runs indicated they did not significantly improve results. They are; So, average basin slope; d, overland flow; and DD, drainage density.

Preliminary Selection of Stations and Their Respective Significant Physical Parameters

The analysis first considered the statistics of 32 gaging stations and thirteen independent variables for each station. It was anticipated that, as the study progressed, some of these stations and most of those parameters would be reduced. This chapter discusses these processes and the final choice of stations, parameters, and regression equations.

A few of the stream gage stations were eliminated immediately as follows; Dry Creek, Pacheco Creek, San Tomas and Wildcat Creek; each had poor data. As previously mentioned, drainage density (DD), overland flow (d) and the average basin slope (So) were immediately eliminated because of their interdependence and the difficulty to compute. The parameter LLS was found to adequately combine all the physical features represented by these parameters.

Regression for the Mean, M

In the initial regression runs, 25 stations and 9 independent variables were used with the following results for the Mean:

RUN #1	R Square	Standard Error
Sm, P, A, Lca, Rb, d/Lm, Ls, Wb, Lm	.880	.244
Sm, P, A, Lca, Rb, d/Lm, Ls, Wb	.880	.237
Sm, P, A, Lca, Rb, d/Lm, Ls	.880	.230
Sm, P, A, Lca, Rb, d/Lm	.880	.224
Sm, P, A, Lca, Rb	.879	.218
Sm, P, A, Lca	.873	.218
Sm, P, A	.835	.242
Sm, P	.780	.274
Sm	.412	.437

Simple Correlation Coefficients

Variable	Area	Map	Rb	Lm	d/Lm
Area	1.0000	0.1362	0.1706	0.9609	-0.3820
Map	0.1362	1.0000	0.0000	0.0900	0.0000
Rb	0.1706	0.0000	1.0000	0.1730	0.0000
Lm	0.9609	0.0900	0.1730	1.0000	-0.4761
d/Lm	-0.3820	0.0000	0.0000	-0.4761	1.0000
Wb	0.9076	0.1330	0.0946	0.7545	-0.1265
Sm	-0.7126	0.0000	-0.1354	-0.6715	0.1122
Ls	0.9841	0.0000	0.2750	0.9550	-0.3704
Lca	0.9254	0.1854	0.2262	0.9679	-0.3994
Mean	0.7556	0.5802	0.0000	0.6765	-0.1762

Variable	Wb	Sm	Ls	Lca	Mean
Area	0.9076	-0.7126	0.9841	0.9254	0.7556
Map	0.1330	0.0000	0.0000	0.1854	0.5802
Rb	0.0946	-0.1354	0.2750	0.2262	0.0000
Lm	0.7545	-0.6715	0.9550	0.9679	0.6765
d/Lm	-0.1265	0.1122	-0.3704	-0.3994	-0.1762
Wb	1.0000	-0.6618	0.8788	0.7185	0.7576
Sm	-0.6618	1.0000	-0.6818	-0.6467	-0.6218
Ls	0.8788	-0.6818	1.0000	0.9220	0.6871
Lca	0.7185	-0.6467	0.9220	1.0000	0.6426
Mean	0.7576	-0.6218	0.6871	0.6426	1.0000

A second trial run eliminated Lca, Lm, and Sm in favor of the combined factor $\frac{Lm\ Lca}{(S_m)^{\frac{1}{2}}}$ (LLS). The results for the Mean, M were:

RUN #2	R Square	Standard Error
Wb, P, LLS, Ls, Rb, d/Lm, A	.823	.280
Wb, P, LLS, Ls, Rb, d/Lm	.823	.272
Wb, P, LLS, Ls, Rb	.821	.265
Wb, P, LLS, Ls	.815	.263
Wb, P, LLS	.799	.268
Wb, P	.777	.276
Wb	.592	.365

It is obvious from these two trials that the shape factor d/Lm is of minor importance and being somewhat difficult to compute was therefore discarded. The high correlation of the four "size" parameters (Lm, Wb, Ls, Lca) with Area caused the reordering in the second trial and the reduction of importance of the single Area factor. In the third trial therefore these four factors were eliminated with the following results:

RUN #3	R Square	Standard Error
Sm, P, A, LLS, Rb	.874	.222
Sm, P, A, LLS	.867	.223
Sm, P, A	.835	.242
Sm, P	.780	.274
Sm	.412	.437

removing LLS as a factor	R Square	Standard Error
Sm, P, A, Rb	.848	.239
Sm, P, A,	.835	.242
Sm, P	.780	.274
Sm	.412	.437

There are a few streamflow stations whose records were considered to be questionable. In an effort to increase the reliability of the regression analysis these were systematically removed until an optimum was reached. The results with 20 stations were:

	R Square	Standard Error
A, P, Sm, Rb	.886	.173
A, P, Sm	.874	.176
A, P,	.842	.191
A	.756	.230

It is obvious from these results that the inclusion of other variables in addition to A and P causes little improvement in the coefficient of determination or the standard error. It was decided, therefore, that the final regression equation of the means should be based on the area and the mean annual precipitation.

Regression for the Standard Deviation

For determination of the independent variables for the Standard Deviation, the same procedure of Runs #1 to #3 as for the Mean was followed. The results for the Standard Deivation were:

RUN #1	R Square	Standard Error
P, Ls, Sm, A, Lm, d/Lm, Lca, Wb, Rb	.456	.123
P, Ls, Sm, A, Lm, d/Lm, Lca, Wb	.455	.119
P, Ls, Sm, A, Lm, d/Lm, Lca	.454	.116
P, Ls, Sm, A, Lm, d/Lm	.452	.113
P, Ls, Sm, A, Lm	.449	.110
P, Ls, Sm, A	.445	.108
P, Ls, Sm	.442	.106
P, Ls	.429	.104
P	.365	.108

RUN #2	R Square	Standard Error
P, Ls, Wb, LLS, d/Lm, A, Rb	.453	.116
P, Ls, Wb, LLS, d/Lm, A	.453	.113
P, Ls, Wb, LLS, d/Lm	.449	.110
P, Ls, Wb, LLS	.448	.108
P, Ls, Wb	.441	.106
P, Ls	.429	.104
P	.365	.108

RUN #3	R Square	Standard Error
P, LLS, Sm, A, Rb	.439	.111
P, LLS, Sm, A	.438	.108
P, LLS, Sm	.437	.106
P, LLS	.420	.105
P	.365	.108

In an effort to increase the reliability of this regression analysis, five streamflow stations with suspect data were removed. The results of this analysis for the remaining 20 stations with only the three parameters A, P, and LLS were:

	R Square	Standard Error
P, LLS, A	.425	.081
P, LLS	.425	.086
P	.263	.078

The final regression equations for the Standard Deviation should therefore, be based on the mean annual precipitation, the basin factor and Area.

Regression for the 1% Peak Directly

A separate regression analysis was made using the 1% peak value as the dependent variable. Much of the District's work requires only the 1% peak value so that a regression of the 1% peak directly might give a better estimate than the mean and standard deviation. Proceeding as for the previous analysis, the results were:

RUN #1	R Square	Standard Error
Ls, Sm, Rb, Lca, P, Lm, Wb, d/Lm, A	.821	.300
Ls, Sm, Rb, Lca, P, Lm, Wb, d/Lm	.820	.290
Ls, Sm, Rb, Lca, P, Lm, Wb	.819	.281
Ls, Sm, Rb, Lca, P, Lm	.817	.273
Ls, Sm, Rb, Lca, P	.811	.269
Ls, Sm, Rb, Lca	.800	.269
Ls, Sm, Rb	.788	.269
Ls, Sm	.761	.279
Ls	.723	.293

RUN #2	R Square	Standard Error
A, Rb, d/Lm, Wb, P, LLS, Ls	.774	.313
A, Rb, d/Lm, Wb, P, LLS	.774	.304
A, Rb, d/Lm, Wb, P	.773	.295
A, Rb, d/Lm, Wb	.772	.287
A, Rb, d/Lm	.770	.281
A, Rb	.763	.278
A	.754	.276

RUN #3	R Square	Standard Error
A, Sm, LLS, Rb, P	.789	.272
A, Sm, LLS, Rb	.783	.269
A, Sm, LLS	.772	.269
A, Sm	.756	.272
A	.735	.277

Removing questionable streamflow data and using the same 20 stations as for the Mean and Standard Deviation the results were:

	R Square	Standard Error
A, Sm, P	.929	.144
A, Sm	.926	.142
A	.917	.146

Therefore, the regression equation for the 1% peak flow should be based on the area, the slope and the mean annual precipitation.

The additional improvement resulting from the use of the Mean Annual Precipitation, P was small. However, the Mean Annual Precipitation was tested separately without Sm, because A & P were the only two factors used in calculating the Mean, M.

	R Square	Standard Error
A, P	.917	.150
A	.917	.146

In each case (i.e. in the case of the mean, standard deviation and 1% peak), the exclusion of several questionable stream gages greatly improved the reliability of the regression analysis. The independent variables also arranged themselves in a much more logical fashion. This improvement illustrates the importance of good data. It is absolutely necessary that each station be carefully scrutinized.

Determination of Final Regression Equations

The primary independent variables input to the regression equations were determined in the preliminary analysis described in the previous section. To further eliminate the impact of questionable data, different combinations of stations were tested separately, adding or excluding stations which were most suspect. During the course of this work, it was noticed that standard errors of regression could be improved by eliminating some stations and adding others.

To the 20 stations used in the previous analysis, three SCVWD stream gages Fisher Creek, Metcalf Creek, and Calabazas Creek were added for a total of 23 stations. This was done in order to stress the applicability of regression to areas similar to those stations, although it is known that the data of these stations is questionable. The following regional equations and their measure of significance resulted:

Equation	R Square	Standard Error
(EQN #1) $M = -.338 + .85 \log A + 1.41 \log p$	.850	.260
(EQN #2) $\sigma = 1.45\ P^{-.42}\ LLS^{.06}$	.468	.074
(EQN #3) $1\% = 19.4\ A^{.95}\ P^{.58}\ Sm^{-.25}$	.904	.218

The Coyote-Madrone station was suspect because of its very high standard deviation, so it was removed from the group of 23 stations and the resulting equations were:

Equation	R Square	Standard Error
(EQN #4) $M = -.300 + .87 \log A + 1.38 \log P$	.841	.267
(EQN #5) $\sigma = 1.29\ P^{-.36}\ LLS^{.04}$	.352	.072
(EQN #6) $1\% = 7.36\ A^{.83}\ P^{.85}\ Sm^{-.33}$	.896	.210

The Zayante Creek station was eliminated next because of its high standard deviation, so that there were only 21 stations in this analysis. The results for these 21 stations were:

Equation	R Square	Standard Error
(EQN #7) $M = -.415 + .85 \log A + 1.47 \log P$	.843	.272
(EQN #8) $\sigma = 1.67 P^{-.46} LLS^{.05}$	.507	.063
(EQN #9) $1\% = 11.19 A^{.86} P^{.73} Sm^{-.30}$	.898	.213

It was very desirable to keep the .8 square mile Metcalf Creek station because the equations are most often applied to small areas. However, there was some question of the reliability of the data, so the station was removed. With the remaining 20 stations the final results were:

Equation	R Square	Standard Error
(EQN #10) $M = .148 + .68 \log A + 1.25 \log P$	.817	.210
(EQN #11) $\sigma = 1.65 P^{-.45} LLS^{.05}$	.502	.065
(EQN #12) $1\% = 238 A^{.84} P^{.23}$	.918	.128

The regression coefficients are tending toward values which generally appear to be reasonable and acceptable. Therefore, further adjustments were not attempted. Equations 10, 11, and 12 have the best combination of multiple determination coefficients and standard errors, and therefore were accepted for the peaks.

Table 5 compares the results of these regression equations with the gage statistics for each of the 20 stations used in the development. The mean and standard deviation for the gage data are the "adopted" statistics from the final HEC Regional Frequency Analysis Program. The -.6 skew is input as a regional constant (this is more fully described in the next chapter).

The other skew values included for comparison are from the recorded and estimated data of the regional frequency program. Similiar analyses were conducted for the volumes and are reviewed in Appendix D of this report. All the resultant equations are summarized in Table 6.

TABLE 5

COMPARISON OF RESULTS OF EQUATIONS

PEAK

Station	Statistics from Recorded and Estimated Data		Equations	
	Adopted	Regional Skew	M (EQN#10) σ (EQN# 11) (g=-.6)	1% (EQN# 2)
1) APTOS				
D.A. = 12.2	M=2.621		M=2.832	
M.A.P. = 36	σ= .536		σ= .436	
LLS = 280.8	g=-.37	g=-.6	1%= 4484	1%=4352
Sm = .0235	1%=5310	1%=4252		
2) BODFISH CREEK				
D.A. = 7.4	M=2.448		M=2.585	
M.A.P.= 30	σ= .488		σ= .438	
LLS = 59.4	g=-.23	g=-.6	1%=2561	1%=2738
Sm = .0327	1%=3177	1%=2319		
3) BRANCIFORTE CREEK				
D.A. = 17.3	M=3.130		M=2.992	
M.A.P. = 40	σ= .357		σ= .419	
LLS = 328.3	g=-.42	g=-.6	1%=6021	1%=5987
Sm = .0172	1%=7156	1%=6326		
4) CALABAZAS CREEK				
D.A. = 4.73	M=2.479		M=2.332	
M.A.P. = 24	σ= .449		σ= .493	
LLS = 83.6	g=-.14	g=-.6	1%=1814	1%=1787
Sm = .0278	1%=3022	1%=2104		
5) CORRALITOS CREEK				
D.A. = 10.6	M=2.615		M=2.848	
M.A.P. = 40	σ= .477		σ= .392	
LLS = 87.8	g=-.30	g=-.6	1%=3845	1%=3954
Sm = .0395	1%=4182	1%=3249		
6) COYOTE CREEK near Gilroy				
D.A. = 109.0	M=3.365		M=3.259	
M.A.P. = 24	σ= .560		σ= .583	
LLS = 2405.0	g=-.77	g=-.6	1%=22,648	1%=25,004
Sm = 0.0141	1%=22,418	1%=26,169		
7) FISHER CREEK				
D.A. = 15.5	M=2.263		M=2.584	
M.A.P. = 20	σ= .663		σ= .591	
LLS = 617.0	g=-.43	g=-.6	1%=4955	1%=4690
Sm = .0025	1%=3941	1%=3231		

TABLE 5 (Continued)

COMPARISON OF RESULTS OF EQUATIONS

PEAK

	Statistics from Recorded and Estimated Data		Equations	
Station	Adopted	Regional Skew	M (EQN#10) σ (EQN# 11) (g=-.6)	1% (EQN# 2)
8) LOS GATOS CREEK				
D.A. = 38.6	M=3.268		M=3.269	
M.A.P. = 43	σ= .421		σ= .415	
LLS = 514.0	g=-1.40	g=-.6	1%=11,199	1%=11,926
Sm = .0260	1%=6663	1%=11,467		
9) MATADERO CREEK				
D.A. = 7.26	M=2.456		M=2.332	
M.A.P. = 19	σ= .469		σ= .566	
LLS = 162.0	g=-.92	g=-.6	1%=2489	1%=2434
Sm = .0138	1%=1716	1%=2176		
10) PERMANENTE CREEK				
D.A. = 15.9	M=2.296		M=2.618	
M.A.P. = 21	σ= .699		σ= .559	
LLS = 309.9	g=-1.47	g=-6	1%=4666	1%=4803
Sm = .0226	1%=1527	1%=4074		
11) REDWOOD CREEK				
D.A. = 1.82	M=2.291		M=1.923	
M.A.P. = 19	σ= .451		σ= .503	
LLS = 15.4	g=-1.09	g=-.6	1%= 739	1%=760
Sm = .0236	1%=967	1%=1376		
12) ROSS CREEK				
D.A. = 2.86	M=2.188		M=2.206	
M.A.P. = 25	σ= .413		σ= .431	
LLS = 8.3	g= -.15	g=-.6	1%=1038	1%=1181
Sm = .0360	1%=1273	1%=921		
13) SAN FRANCISQUITO				
D.A. = 37.5	M=3.044		M=3.083	
M.A.P. = 31	σ= .551		σ= .485	
LLS = 606.7	g=-1.03	g=-.6	1%=9,880	1%=10,826
Sm = .0129	1%=8000	1%=12,019		

TABLE 5 (Continued)

COMPARISON OF RESULTS OF EQUATIONS

PEAK

Station	Statistics from Recorded and Estimated Data		Equations	
	Adopted	Regional Skew	M (EQN#10) σ (EQN# 11) (g=-.6)	1% (EQN# 2)
14) SAN LORENZO CREEK at Hayward				
D.A. = 37.5	M=2.987		M=2.944	
M.A.P. = 24	σ= .579		σ= .527	
LLS = 324.1	g=-.72	g=-.6	1%=8605	1%=10,200
Sm = .0152	1%=10,695	1%=11,899		
15) SAN LORENZO RIVER				
D.A. = 111.0	M=3.726		M=3.640	
M.A.P. = 48	σ= .400		σ= .428	
LLS = 2519.6	g=-.57	g=-.6	1%=27,838	1%=29,860
Sm = .0071	1%=31,189	1%=30,060		
16) SARATOGA CREEK				
D.A. = 9.2	M=2.598		M=2.832	
M.A.P. = 42	σ= .364		σ= .378	
LLS = 64.9	g=+.06	g=-.6	1%=3488	1%=3547
Sm = 0.0500	1%=2888	1%=1915		
17) SOQUEL CREEK				
D.A. = 40.2	M=3.295		M=3.241	
M.A.P. = 40	σ= .411		σ= .446	
LLS =1156.2	g=-.53	g=-.6	1%=12,008	1%=12,163
Sm = .015	1%=12,370	1%=11,686		
18) UVAS CREEK at Morgan Hill				
D.A. = 30.3	M=3.391		M=3.121	
M.A.P. = 37.4	σ= .348		σ= .437	
LLS = 407.6	g=-.36	g=-.6	1%=8761	1%=9432
Sm = 0.0207	1%=12,820	1%=11,097		
19) UVAS CREEK above Reservoir				
D.A. = 21.0	M=3.318		M=3.036	
M.A.P. = 39	σ= .364		σ= .416	
LLS = 220.5	g=-.62	g=-.6	1%=6578	1%=6996
Sm = .0234	1%=9970	1%=10,053		
20) WEST BRANCH SOQUEL CREEK				
D.A. = 12.2	M=2.974		M=2.953	
M.A.P. = 45	σ= .471		σ= .392	
LLS = 246.8	g=-.58	g=-.6	1%=4897	1%=4582
Sm = .0224	1%=7394	1%=7235		

TABLE 6

SUMMARY
OF REGRESSION EQUATIONS

	R Square	Standard Error
PEAK (cfs)		
$M = .148 + .68 \log A + 1.25 \log P$	.817	.210
$\sigma = 1.65\ P^{-.45}\ LLS^{.05}$	.502	.065
$1\% = 238\ A^{.84}\ P^{.23}$	.918	.128
24-HR Volume (average cfs)		
$M = -.756 + .83 \log A + 1.51 \log P$	.891	.188
$\sigma = 1.805\ P^{-.46}\ LLS^{.04}$	.514	.063
$1\% = 11.625\ A^{.90}\ P^{.64}\ S_m^{-.15}$	.958	.111
1-Day Volume (1-Day average cfs)		
$M = -1.039 + .88 \log A + 1.60 \log P$	.904	.186
$\sigma = 1.752\ P^{-.46}\ LLS^{.05}$	.462	.063
$1\% = 8.558\ A^{.98}\ P^{.67}\ S_m^{-.10}$	.969	.100
2-Day Volume (Sum of 2-Day Average cfs)		
$M = -.898 + .88 \log A + 1.61 \log P$	.901	.179
$\sigma = 1.872\ P^{-.47}\ LLS^{.04}$	.51	.064
$1\% = 11.165\ A^{.96}\ P^{.69}\ S_m^{-.12}$	.965	.097
3-Day Volume (Sum of 3-Day Average cfs)		
$M = -.859 + .88 \log A + 1.64 \log P$	.918	.173
$\sigma = 1.949\ P^{-.48}\ LLS^{.04}$	.490	.068
$1\% = 10.206\ A^{.95}\ P^{.75}\ S_m^{-.15}$	.967	.102

A. D. Saah

Small Basins

In use, the regional regression equations would be applied to small tributary basins and then combined with adjacent basins and routed until the entire watershed has been included. Since the application would be mostly for smaller basins the 1% regression was tested excluding all stream gages with greater than 30 square miles of drainage. The results with 14 stations are:

	R Square	Standard Error
A, P, Sm	.804	.162
A, P	.802	.155
A	.782	.155

the regression equation is:

(EQN #13) $Q_{1\%} = 175\ A^{.76}\ P^{.36}$

This equation should be used for checking only.

DETERMINATION OF SKEW COEFFICIENT

Previously, the discussion included the determination of the two statistics, mean and standard deviation. The third statistic, needed to fully define a frequency curve, is the coefficient of skewness. This statistic is sensitive to the number of years of record and hence more difficult to calculate. Matalas, Nicholas and Benson found that the standard error of the skew coefficient can go as high as 0.45 with 25 years of record. This standard error value increases to 0.70 with only 10 years of record and drops to only 0.25 with 100 years of record. For fifty years of record, the standard error is 0.35. In other words, for small samples the dispersion of the calculated skew coefficient around the skew obtained from the population prohibits casual application of this parameter.

For this reason and because the average length of records in this area is less than 25 years, it was decided to use a regionalized skew coefficient where a significant improvement on the standard error could be achieved. This regional skew was determined by computing the arithmetic and weighted means of the skew factors for the recorded, and recorded and estimated data. Table 7 shows these values for various durations. The final regional skew factors selected for future studies were -0.6, -0.64, -0.66, -0.68, -0.69, -0.70, -0.72, and -0.74 for the peak, 1-day, 2-day, 3-day, 4-day, 5-day, 10-day, and 15-day volumes, respectively.

These values match the framework of the regional skew factors as outlined by the W.R.C. It was contemplated at one stage of this study that isolines of equal skews in the District would

TABLE 7

REGIONAL FREQUENCY ANALYSIS

Skewness Coeff. Statistics Criteria	FROM RECORDED DATA ALONE	
	$\bar{g}$	$\bar{g}_w$
Peak	-0.991	-0.980
1-Day	-0.799	-0.789
2-Day	-0.834	-0.813
3-Day	-0.857	-0.850
4-Day	-0.863	-0.861
5-Day	-0.844	-0.848
10-Day	-0.840	-0.858
15-Day	-0.827	-0.856

Skewness Coeff. Statistics Criteria	FROM RECORDED AND ESTIMATED DATA	
	$\bar{g}$	$\bar{g}_w$
Peak	-0.600	-0.612
1-Day	-0.640	-0.643
2-Day	-0.630	-0.664
3-Day	-0.680	-0.679
4-Day	-0.690	-0.687
5-Day	-0.700	-0.698
10-Day	-0.720	-0.723
15-Day	-0.740	-0.738

$\bar{g}$ - Arithmetic Mean

$\bar{g}_w$ - Weighted Mean

be valuable to the determination of a regional skew factor. But after the calculated skew coefficients were plotted, a trend was not definable at all. So it was decided that until more data points are available, such an attempt could yield misleading results.

The regional information should be supplemented by site specific information. Therefore the W.R.C. method of obtaining an adopted skew by weighing the regional skew with the skew obtained from recorded will be utilized.

Due to the difficulty of determining skew with the existing records the next updating of the regression equations should concentrate on the skew coefficient.

The following Table 8 shows the calculation of the various skew factors. A close look at the table reveals that the most likely skewness to include most of the information is the skew obtained from the averaged adopted skews (-.59 or -.60). This skew factor satisfied the requirement of the W.R.C. as applied to gaged locations and satisfies the criteria of using a weighted skew coefficient rather than the calculated one.

TABLE 8

Adopted Regional Coefficient of Skewness
(Calculated Regional Coefficient of Skew = -.60)

	Recorded and Estimated Data			Recorded Data Only		
	Skew	Ne	Adopted g *	Skew	N	Adopted g_w*
Aptos	-.366	58.3	-.50	-1.125	17	-.6
Bodfish	-.230	54.0	-.46	-.411	16	-.6
Branciforte	-.419	52.6	-.53	-.230	19	-.6
Calabazas	-.136	55.1	-.41	+.138	35	-.50
Corralitos	-.303	56.9	-.47	-.920	15	-.6
Coyote-Gilroy	-.767	58.7	-.68	-1.676	15	-.6
Fisher	-.432	55.7	-.53	-.360	23	-.6
Los Gatos	-.139	55.5	-.41	-.566	15	-.6
Matadero	-.923	56.8	-.74	-1.099	22	-.6
Permanente	-1.470	57.8	-.98	-1.902	23	-.6
Redwood	-1.085	58.6	-.82	-1.610	16	-.6
Ross	-.152	51.7	-.44	-.427	33	-.58
San Francisquito	-1.030	53.5	-.76	-1.805	35	-.76
San Lorenzo Creek	-.717	53.3	-.64	-.456	29	-.59
San Lorenzo River	-.572	55.7	-.59	-.348	23	-.6
Saratoga	.057	52.2	-.36	+.254	42	-.41
Soquel	-.530	47.9	-.58	-.970	24	-.6
Uvas MH	-.359	56.1	-.50	-.392	27	-.59
Uvas Reservoir	-.619	41.4	-.60	-1.579	14	-.6
West Br. Soquel	-.575	60.3	-.59	-.820	14	-.6
		Weighted =	-.58	Weighted =		-.59

*Weighted skew, $g_w = \frac{(N-25)}{75}$ Recorded + $(1 - \frac{N-25}{75})$ Regional (-.6)

Where N $\geqslant$ 25 years.

CONFIDENCE LIMITS

The accuracy of results obtained from equations derived from hydrologic data are limited by errors inherent in the data itself, and by the choice of a distribution function that defines the frequency of that data. Therefore the equations should be complemented by an associated standard error function, or an index of reliability called the confidence band. Similar to the hydrologic processes from which the equations were derived, the standard errors of the dependent variables of the equations (mean and standard deviation), follow certain distributions.

In the case of the mean and standard deviation, the best suited distributions are the student "t" and the chi-square respectively. The choice of deviates from these distributions depends on the choice of a probability of occurrence. This probability is then called the confidence limit of our statistical results.

An 80% confidence limit or the 80% level of significance means that, regardless of whatever sample is chosen from the same population, there is an 80% certainty that the resultant value will be equal to or less than the calculated upper confidence limit, and equal to or more than the calculated lower confidence limit. The upper 80% limit could also be called the upper 10% confidence limit, which is similar in that there is only a 10% chance that the value will be greater than the calculated upper 10% confidence limit value.

Utilizing the known characteristics of the student "t" and chi-square distributions, the standard error of the means ε_m (standard deviation of the means of many samples from the same population) is calculated by the following formula:

$$\varepsilon_m = \frac{\sigma^2}{\sqrt{N_e}}$$

Where σ = Standard deviation.

N_e = Number of years.

Similarly the standard error of the standard deviation ε_σ (standard deviation of the standard deviations of many samples from the same population) is calculated as follows:

$$\varepsilon_\sigma = \frac{\sigma^2}{\sqrt{2\ N_e}}$$

The combined error of the mean and standard deviation is distributed according to a non-central "t" distribution which was adopted and used to adjust the already adopted frequency factor (k) as per the following formula:

$$Q = M + k\ (^{Kc}/K)\ \sigma$$

Where $(^{Kc}/K)$ is the adjusting ratio from a normal distribution (k).

As presented in the W.R.C. Bulletin #17, the same non-central "t" distribution is used and tables of Kc/K are available. The question is what are the equivalent years of record (Ne) that should be attributed to the regression results to determine the non-central deviates?

The number of gaging stations and their years of record are both factors which should be used in defining the equivalent number of years (station-years) used for the determination of a confidence band.

Assuming that annual flood series are correlated among stations, the equivalent number of stations, n_e, is given by the following formula:

$$n_e = \frac{n}{1 + \bar{r}\ (n-1)}$$

n = number of stations

$\bar{r}$ = average correlation coefficient of all correlation coefficients for a matrix of $n\ (n-1)\ /2$ different values.

The equivalent number of station-years is, therefore, equal to:

$$N_e = No(n_e) = \frac{No(n)}{1 + \bar{r}\ (n-1)}$$

where;

No = the average number of years of record of n stations.

From the present correlation studies the average correlation coefficient $\bar{r} = .785$ and the number of stations used in the study was 20. Therefore, the equivalent number of stations (n_e) for the regional study is equal to:

$$n_e = \frac{20}{1 + .785\ (19)} \qquad = 1.256$$

Knowing the average number of equivalent years in each station, No = 55, the equivalent number of station-years, therefore, would be equal to:

$$N_e = No \times n_e$$

$$= 55 \times 1.256$$

$$= 69.08$$ Say 69 years

Table 9 shows the appropriate ratios to be multiplied by the calculated deviates of a near normal frequency distribution to obtain the non-central deviates of a preassigned confidence band probability. These could be used in the computation of the upper 80% (10%) confidence band as obtained from the W.R.C. Bulletin #17 (the sample size N_e = 69 years).

TABLE 9*

Probability	k_c/k
.001	1.14
.005	1.14
.010	1.14
.020	1.15
.050	1.16
.100	1.18
.200	1.23
.400	1.65

k_c could then be calculated if the deviate is obtained from a near normal distribution.

---

*This table was adapted from the W.R.C. Bulletin #17.

APPLICATION OF REGIONAL EQUATIONS

General

The formulas can be applied to gaged and ungaged watersheds. It should be strongly emphasized that all available information must be evaluated before a decision on a final flow rate is made. Careless application of the equations could produce misleading results. Hydrology is after all as much an art as a science and numerical results must be thoroughly backed with experience and judgment. For ungaged watersheds the mean and standard deviation obtained from regression equations are considered to contain the best available information and are used directly. For gaged watersheds though, the site specific results of the frequency analysis of recorded data could be weighted with the results from the regional equations. The weights to be used are those of the respective equivalent years of record obtained from the standard errors of the various statistics used, as given below.

Equivalent Years of Record

For the regression equations of the peaks, the standard errors of the mean and standard deviation are numerically known (Table 6). Hence the equivalent years of record for the mean are obtained from these standard errors as calculated by the following formula:

$$N_{e,m} = \frac{\sigma^{2}}{\varepsilon_{m}^{2}}$$

Where ε_m is the standard error of mean

σ is the standard deviation

$N_{e,m}$ is equivalent years of record of the mean

Similarily, the equivalent years of record of the standard deviation

$$N_{e,\sigma} = \frac{\sigma^2}{2\left(\varepsilon_\sigma\right)^2}$$

N_e is then multiplied by its own statistic. The statistics of recorded data is similarly multiplied by the number of years of record (N). The results are then added and divided by the summation of the N_e and N.

Consequently, the weighted mean of regional and recorded results is:

$$M = \frac{M_r \cdot N_{e,m} + M_s \cdot N}{N_{e,m} + N}$$

Where M is the weighted mean

M_r is the mean obtained from regression

M_s is the mean obtained from recorded data

Similarly,

$$\sigma^2 = \frac{\sigma_r^2 \cdot N_{e,\sigma} + \sigma_s^2 \cdot N}{N_{e,\sigma} + N}$$

Where σ, σ_r and σ_s are the weighted standard deviation, the one obtained from regression and one from recorded data respectively.

Sensitivity Analysis

The values resulting from the application of the regional regression equations at the 20 base stations together with the statistical values of the recorded date are summarized in Table 10.

TABLE 10

1976 REGIONAL EQUATIONS

	RECORDED DATA				SE, Mean = .210			SE,o = .065		WEIGHTED VALUES			Base Weighted Values				
STATION	Mean	σ	g	N	Mean	σ	g	NE Mean	NE σ	Mean	σ	g	1%	50%	1976 Regional 1%	1976 Regional M,σ	U.S.G.S. Graphically
Aptos	2.616	.583	-1.125	17	2.832	.436	-.6	4.3	22.5	2.660	.499	-.6	3,964	512	4,352	4,484	4,400
Bodfish	2.445	.479	- .411	16	2.585	.438	-.6	4.4	22.7	2.475	.455	-.6	2,140	331	2,738	2,561	2,400
Branciforte	3.176	.381	- .230	19	2.992	.419	-.6	4.0	20.8	3.144	.401	-.6	7,904	1527	5,987	6,021	6,000
Calabazas	2.485	.431	+ .138	35	2.332	.493	-.6	5.5	28.8	2.464	.459	-.50	2,298		1,787	1,814	1,700
Corralitos	2.639	.538	- .920	15	2.848	.392	-.6	3.5	18.2	2.679	.458	-.6	3,468	530	3,954	3,845	4,200
Coyote-Gilroy	3.334	.612	-1.676	15	3.259	.583	-.6	7.7	40.2	3.309	.591	-.6	26,307	2331	25,004	22,648	22,000
Fisher	2.337	.523	- .360	23	2.584	.591	-.6	7.9	41.3	2.400	.567	-.6	2,923	286	4,690	4,955	4,700
Los Gatos	3.261	.441	- .566	15	3.269	.415	-.6	3.9	20.4	3.263	.426	-.6	11,585	2019	11,926	11,199	12,000
Matadero	2.445	.455	-1.099	22	2.332	.566	-.6	7.3	37.9	2.417	.525	-.6	2,535	294	2,434	2,489	2,500
Permanente	2.255	.767	-1.902	23	2.618	.559	-.6	7.1	37.0	2.341	.633	-.6	3,486	254	4,803	4,666	4,700
Redwood	2.310	.403	-1.610	16	1.923	.503	-.6	5.7	29.9	2.208	.468	-.6	1,224	180	760	739	570
Ross	2.195	.392	- .427	33	2.206	.431	-.6	4.2	22.0	2.196	.408	-.58	931		1,181	1,038	1,200
San Francisquito	3.155	.531	-1.805	35	3.083	.485	-.6	5.3	27.8	3.146	.511	-.76	11,127		10,826	9,880	10,000
San Lorenzo Cr.	2.970	.548	- .456	29	2.944	.527	-.6	6.3	32.9	2.965	.537	-.59	9,520		10,200	8,605	9,500
San Lorenzo Riv.	3.778	.452	- .348	23	3.640	.428	-.6	4.2	21.7	3.757	.440	-.6	38,388	6318	29,860	27,838	29,000
Saratoga	2.621	.379	+ .254	42	2.832	.378	-.6	3.2	16.9	2.636	.379	-.41	2,525		3,547	3,488	3,800
Soquel	3.345	.461	- .970	24	3.241	.446	-.6	4.5	23.5	3.329	.454	-.6	15,224	2366	12,163	12,008	12,000
Uvas MH	3.375	.402	- .392	27	3.121	.437	-.6	4.3	22.6	3.340	.418	-.59	13,464		9,432	8,761	9,000
Uvas Res.	3.384	.303	-1.579	14	3.036	.416	-.6	3.9	20.5	3.308	.370	-.6	10,083	2211	6,996	6,578	6,800
West Br. Soquel	2.956	.532	- .820	14	2.953	.392	-.6	3.5	18.2	2.955	.453	-.6	6,407	1000	4,582	4,897	4,600

$NE_{Mean} = \frac{\sigma^2}{(SE)^2}$ $NE_o = \frac{\sigma^2}{2(SE)^2}$ $g = \frac{(N-25)}{75}$ Recorded + $(1-\frac{N-25)}{75}$ Regional

$NE_{Mean} = \frac{\sigma^2}{(.210)^2}$ $= \frac{\sigma^2}{2(.065)^2}$

$\pm\Delta$ = 18,281	$\pm\Delta$ = 26,989	$\pm\Delta$ = 27,352	$\pm\Delta$ = - 24,433
$\lvert\Delta\rvert$ = 31,979	$\lvert\Delta\rvert$ = 38,189	$\lvert\Delta\rvert$ = 38,428	$\lvert\Delta\rvert$ = 37,189
Δ^2 = 124,450,427	Δ^2 = 185,325,299	Δ^2 = 216,647,948	Δ^2 = 164,027,149

The best estimate of the desired flowrates is the weighted average of the value obtained from the correlation and extension of data and the value computed from the regional equations. The applicability of the various procedural possibilities is measured by the least square of the deviations from those average values.

1976 Regional Eqn. (1%)	$\Delta^2 = 7.7 \times 10^6$
U.S.G.S. Eqn.	$\Delta^2 = 25.5 \times 10^6$
1976 Regional Eqn. (M_σ)	$\Delta^2 = 29.1 \times 10^6$
Weighted Regional and Recorded	$\Delta^2 = 103.6 \times 10^6$
1972 Regional Eqn.	$\Delta^2 = 519.0 \times 10^6$

It appears that in general the 1976 Regional Equations should be preferred. However, there are, most certainly, specific cases where this may not be true. Appendix E is a graphical representation of those results.

Volumes Equations

Designation of design floods includes the designation of peaks and their related volumes. A regression analysis study was done to determine the 24 Hr., 1-day, 2-day, and 3-day volumes and their frequency of occurrence. This study is explained in detail in Appendix D.

To combine a flood peak that has an assigned frequency of occurrence with a flood volume that has another frequency of occurrence creates problems associated with the compounded probabilities as per the following formula.

$$P_{(P,V/_P)} = P_{(P)} \times P_{(V/_P)} = P\ (F)$$

Where $P_{(P,V/_P)}$ is the probability of the peak and volumes occurring together

$P_{(P)}$ is the probability of the peak

$P_{(V/p)}$ is the probability of the volume given the peaks.

The value of $P\,(V/p)$ would be around unity if the volumes associated with the peaks are selected.

Range of Application

The areas of gaged basins used for the analysis range from 0.8 square mile to 196 square miles and the basin average mean-annual precipitation ranging from 17 inches to 50 inches. This is an indication of the range of validity of the regional equations.

From studying the distribution of the M.A.P.'s and Areas, it appears that the most applicable range is less than 50 square miles of area. Experience with small basins indicates that additional care is needed when small basins are encountered. Applicability of the mean annual precipitation is uniform for the full range of values encountered.

In summary, the regression equations are best applied with the following limitations:

1. Mean annual precipitation ranges from 12" to 56".
2. The elevation range is from sea level to an elevation of 4000 feet, central pacific coast.
3. Average slope of main channel ranges from .002 to .05 ft./ft.
4. Areas ranging from 1 to 200 square miles although for areas less than 30 square miles, the results should be checked with EQN #13.

REFERENCES

1. Leo R. Beard, "Statistical Methods in Hydrology", January, 1962. (Rev. October, 1964).
2. "Methods of Flow Frequency Analysis", prepared under the auspices of the subcommittee on hydrology, Inter-Agency Committee on Water Resources, Bulletin #13, April, 1966.
3. H. C. Riggs, "Some Statistical Tools in Hydrology", U.S.G.S.
4. "Regional Frequency Computations", Hydrologic Engineering Center Program 723-X6-L2350, together with the 1972 additions.
5. "Guide to Hydrometeorological Practices", Work Meteorological Organization (WMO-No. 168.TP.82), Switzerland, 1970.
6. "Guidelines for Flood Flow Frequency", W.R.C. Bulletin #17, March, 1976.
7. V. T. Chow, "Handbook of Applied Hydrology", 1964.
8. J. E. Nash et.al., "Flood Frequency as a Function of Catchment Characteristics", Proc. of I.C.E. (London) Symposium, 1966.
9. Langbein, W. B., "Annual Floods and the Partial Duration Series", Tran. W.G.W., Volume 30, 1949.
10. Matalas, N. and B. Jacogs, "A Corrolation Procedure for Augmenting Hydrologic Data",U.S.G.S. Pro. Paper #434E, 1964.
11. Myron B. Feiring,"Use of Corrolation to Improve Estimates of the Mean and Variance", U.S.G.S. Professional Paper #434-C.
12. Beard, Dawdy, Grigg, James & Yevjevich, "Review of Basic Hydrology Methodology for Flood Control", S.C.V.W.D., 1976.
13. A. Saah, "Summary, Basic Hydrology Methods", S.C.V.W.D., 1975.
14. A. Saah, "San Francisquito Creek Hydrology", S.C.V.W.D., Unpublished Report, 1976.
15. "Multiple Linear Regression", The Hydrologic Engineering Center Generalized Computer Program 704-G1-L202.
16. N. Tvwonivk and K. Wiebe, "Application of Regression Analysis in Hydrology", Technical Bulletin #24.
17. M. E. Jennings and M. A. Benson, "Frequency Curves for Annual Flood Series With Incomplete Data", U.S.G.S., February, 1969.

APPENDIX D*

Volumes Equations

---

* Appendexes A, B, and C are listed under the Table of Contents and are available upon request.

Regression for 24-Hour Volume

Maximum annual 1-day streamflows (usually corresponding to maximum annual peak) were extracted from the data and converted to 24-hour volumes through the relationship

$$\log_{24} = .23 \log Q_{peak} + .77 \log Q_{daily}$$

developed by the Corps of Engineers. The dependent variables (mean, standard deviation and 1% flow (g = -.64) of this recorded and reconstituted 24-hour data were calculated and listed in Table 1.

The same 20 streamflow stations and 4 independent variables used in the analysis for the peak were used for the 24-hour volume.

The final resulting regression equations for the 24-hour volumes are:

	R Square	Standard Error
$M = -.756 + .83 \log A + 1.51 \log P$	.891	.188
$\sigma = 1.805\ P^{-.46}\ LLS^{.04}$	.514	.063
$1\% = 11.625\ A^{.90}\ P^{.64}\ S_m^{-.15}$	.958	.111

The results of the application of these equations are compared with the results of the statistical analysis of recorded data at same location. Table 2 shows this comparison.

TABLE 1

REGIONAL REGRESSION ANALYSIS
STREAM GAGE STATISTICS
24-HOUR VOLUME

	Mean	Standard Deviation	1% 24-Hr. Volume Avg. cfs (g = -.64)
Aptos Creek*	2.2944	.5062	1701
Bodfish Creek*	2.0698	.5221	1086
Branciforte Creek*	2.8398	.3562	3153
Calabazas Creek* SCVWD #31	2.0109	.4083	584
Cedar Creek	2.0676	.8326	4054
Corralitos Creek*	2.3573	.4714	1696
Coyote Creek* near Gilroy	3.1455	.5344	13619
Coyote Creek at Madrone	3.2054	.7603	40922
Dry Creek	1.8265	.7370	1549
Fisher Creek* SCVWD #10	1.9832	.7158	2030
Hale Creek SCVWD #33	1.5157	.6124	445
Los Animas Creek SCVWD #56	2.0411	.7186	2347
Los Gatos Creek*	3.0709	.4106	6769
Matadero Creek*	1.9820	.4823	749
Metcalf Creek SCVWD #40	.3890	.3328	10
Pacheco Creek	2.7837	.9588	36097
Packwood Creek SCVWD #57	1.3202	.7330	474
Permanente Creek* SCVWD #34	1.8783	.6542	1226
Redwood Creek*	1.7585	.4554	399

---

*The final 20 stations used for the peak regression analysis.

TABLE 1

REGIONAL REGRESSION ANALYSIS (continued)
STREAM GAGE STATISTICS
24-HOUR VOLUME

	Mean	Standard Deviation	1% 24-Hr. Volume Avg. cfs (g = -.64)
Ross Creek* SCVWD #21	1.6695	.4587	330
San Francisquito*	2.7346	.5509	5672
San Lorenzo Creek* at Hayward	2.6553	.5718	5166
San Lorenzo River*	3.5286	.3927	17992
San Tomas Creek SCVWD #29	1.7655	.6374	880
Saratoga Creek* at Saratoga	2.3225	.3506	936
Soquel Creek*	3.0438	.4052	6215
Upper Permanente SCVWD #32	1.5518	.3771	178
Uvas Creek* at Morgan Hill	3.1012	.3579	5798
Uvas Creek* above Uvas Reservoir	2.9440	.3974	4777
West Branch Soquel Creek*	2.6144	.4505	2804
Wildcat Creek SCVWD #30	2.0497	.4211	674
Zayante Creek	2.5392	.4974	2880

---

*The final 20 stations used for the peak regression analysis.

TABLE 2
COMPARISON OF RESULTS OF PRELIMINARY EQUATIONS
24-HOUR VOLUME

	Station	Gage* Statistics Avg. cfs	1976 STUDY 24 HR. AVG. CFS. M.σ (g=-.64)	1976 STUDY 24 HR. AVG. CFS. 1%
1)	APTOS CREEK	M=2.294 σ= .506 1%=1701	M=2.496 σ= .435 1%= 1997	1%=1920
2)	BODFISH CREEK	M=2.070 σ= .522 1%=1086	M=2.196 σ= .445 1%= 1043	1%=1037
3)	BRANCIFORTE CREEK	M=2.840 σ= .356 1%=3153	M=2.691 σ= .417 1%= 2899	1%=2949
4)	CALABAZAS CREEK SCVWD #31	M=2.011 σ= .408 1%=584	M=1.888 σ= .499 1%= 649	1%=616
5)	CORRALITOS CREEK	M=2.357 σ= .471 1%=1696	M=2.514 σ= .396 1%= 1762	1%=1675
6)	COYOTE CREEK near Gilroy	M=3.146 σ= .534 1%=13619	M=3.019 σ= .571 1%=11,912	1%=11,482
7)	FISHER CREEK SCVWD #10	M=1.983 σ= .716 1%=2030	M=2.197 σ= .588 1%= 1927	1%=2289
8)	LOS GATOS CREEK	M=3.071 σ= .411 1%=6769	M=3.027 σ= .411 1%= 6126	1%=5977
9)	MATADERO CREEK	M=1.982 σ= .482 1%=749	M=1.889 σ= .571 1%= 883	1%=866
10)	PERMANENTE CREEK SCVWD #34	M=1.878 σ= .654 1%=1226	M=2.238 σ= .560 1%= 1875	1%=1737

*24-Hour flows (from 1-day conversion $\log Q_{24} = .23 \log Q_{peak} + .77 \log Q_{daily}$)

TABLE 2 (continued
COMPARISON OF RESULTS OF PRELIMINARY EQUATIONS
24-HOUR VOLUME

	Station	Gage* Statistics Avg. cfs	1976 STUDY 24 HR. AVG. CFS.	
			M.σ (g=-.64)	1%
11)	REDWOOD CREEK	M=1.759 σ= .455 1%=399	M=1.391 σ= .520 1%= 225	1%=230
12)	ROSS CREEK SCVWD #21	M=1.670 σ= .459 1%=330	M=1.734 σ= .447 1%= 363	1%=387
13)	SAN FRANCISQUITO CREEK	M=2.735 σ= .551 1%=5672	M=2.802 σ= .481 1%= 4914	1%=5247
14)	SAN LORENZO CREEK at Hayward	M=2.655 σ= .572 1%=5166	M=2.635 σ= .527 1%= 4073	1%=4346
15)	SAN LORENZO RIVER	M=3.529 σ= .393 1%=17992	M=3.480 σ= .416 1%=17,783	1%=20,160
16)	SARATOGA CREEK at Saratoga	M=2.323 σ= .351 1%=936	M=2.495 σ= .382 1%=1593	1%=1468
17)	SOQUEL CREEK	M=3.044 σ= .405 1%=6215	M=2.995 σ= .439 1%=6397	1%=6428
18)	UVAS CREEK at Morgan Hill	M=3.101 σ= .358 1%=5798	M=2.849 σ= .434 1%= 4480	1%=4549
19)	UVAS CREEK above Reservoir	M=2.944 σ= .397 1%=4777	M=2.744 σ= .415 1%= 3252	1%=3298
20)	WEST BRANCH SOQUEL CREEK	M=2.614 σ= .451 1%=2804	M=2.642 σ= .391 1%= 2315	1%=2232

*24-Hour flows (from 1-day conversion $\log Q_{24} = .23 \log Q_{peak} + .77 \log Q_{daily}$)

1-Day, 2-Day, and 3-Day Volumes

Tables 3, 4 and 5 are the statistics obtained from recorded and estimated data for the maximum 1-day, 2-day, and 3-day volumes. These are the dependent variables used in the regression.

The independent variables are the same used for the peaks and 24-hour volumes.

A summary of the resultant regression equations and their level of significance are shown in Table 6 of the main text.

TABLE 3

REGIONAL REGRESSION ANALYSIS
STREAM GAGE STATISTICS
1-DAY AVERAGE CFS

	Mean	Standard Deviation	1% 1-Day Volume (g = -.64)**
Aptos Creek*	2.197	.508	1370
Bodfish Creek*	1.957	.505	778
Branciforte Creek*	2.753	.352	2536
Calabazas Creek* SCVWD #31	1.872	.436	477
Cedar Creek	1.980	.837	3376
Corralitos Creek*	2.280	.470	1410
Coyote Creek* near Gilroy	3.080	.544	12201
Coyote Creek at Madrone	3.124	.768	35059
Dry Creek	1.714	.736	1190
Fisher Creek* SCVWD #10	1.902	.684	1470
Hale Creek SCVWD #33	1.320	.799	628
Los Animas Creek SCVWD #56	1.953	.720	1927
Los Gatos Creek*	3.012	.414	5996
Matadero Creek*	1.841	.472	518
Pacheco Creek	2.757	.911	27692
Packwood Creek SCVWD #57	1.180	.833	526
Permanente Creek* SCVWD #34	1.737	.707	1109
Redwood Creek*	1.597	.447	265

---

*The final 20 stations used for the peak regression analysis.

**For skew g = -.64 K = 1.850

TABLE 3 (Continued)

REGIONAL REGRESSION ANALYSIS
STREAM GAGE STATISTICS
1-DAY AVERAGE CFS

	Mean	Standard Deviation	1% 1-Day Volume (g = -.64)**
Ross Creek* SCVWD #21	1.516	.452	225
San Francisquito*	2.642	.561	4785
San Lorenzo Creek* at Hayward	2.556	.566	4009
San Lorenzo River*	3.470	.397	16012
San Tomas Creek SCVWD #29	1.623	.702	835
Saratoga Creek* at Saratoga	2.240	.358	798
Soquel Creek*	2.969	.417	5501
Upper Permanente SCVWD #32	1.455	.390	150
Uvas Creek* at Morgan Hill	3.015	.359	4777
Uvas Creek* above Uvas Reservoir	2.832	.397	3685
West Branch Soquel Creek*	2.507	.456	2242
Wildcat Creek SCVWD #30	1.915	.438	531
Zayante Creek	2.454	.513	2529

---

*The final 20 stations used for the peak regression analysis

**For skew g = .64   K = 1.850

TABLE 4

REGIONAL REGRESSION ANALYSIS
STREAM GAGE STATISTICS
SUM OF 2-DAY AVERAGE CFS

	Mean	Standard Deviation	1% 2-Day Volume (g = -.66)**
Aptos Creek*	2.372	.474	1747
Bodfish Creek*	2.124	.522	1208
Branciforte Creek*	2.917	.344	3536
Calabazas Creek* SCVWD #31	2.044	.425	667
Cedar Creek	2.156	.755	3484
Corralitos Creek*	2.460	.461	2025
Coyote Creek* near Gilroy	3.278	.516	16802
Coyote Creek at Madrone	3.315	.710	41549
Dry Creek	1.883	.703	1492
Fisher Creek* SCVWD #10	2.070	.715	2414
Hale Creek SCVWD #33	1.474	.808	907
Los Animas Creek SCVWD #56	2.109	.709	2575
Los Gatos Creek*	3.190	.398	8332
Matadero Creek*	1.989	.474	723
Pacheco Creek	2.937	.898	38523
Packwood Creek SCVWD #57	1.377	.752	572
Permanente Creek* SCVWD #34	1.908	.714	1655
Redwood Creek*	1.736	.444	356

---

*The final 20 stations used for the peak regression analysis.

**For skew g = -.66 K = 1.836

TABLE 4 (Continued)

REGIONAL REGRESSION ANALYSIS
STREAM GAGE STATISTICS
SUM OF 2-DAY AVERAGE CFS

	Mean	Standard Deviation	1% 2-Day Volume (g = -.66)**
Ross Creek* SCVWD #21	1.700	.481	383
San Francisquito*	2.813	.573	7329
San Lorenzo Creek* at Hayward	2.721	.550	5380
San Lorenzo River*	3.636	.391	22588
San Tomas Creek SCVWD #29	1.811	.690	1196
Saratoga Creek* at Saratoga	2.443	.350	1217
Soquel Creek*	3.126	.428	8162
Upper Permanente SCVWD #32	1.677	.451	320
Uvas Creek* at Morgan Hill	3.181	.375	7404
Uvas Creek* above Uvas Reservoir	2.977	.434	5940
West Branch Soquel Creek*	2.672	.440	3019
Wildcat Creek SCVWD #30	2.085	.423	727
Zayante Creek	2.587	.464	2747

---

*The final 20 stations used for the peak regression analysis.

**For skew g = -.66 K = 1.836

TABLE 5

REGIONAL REGRESSION ANALYSIS
STREAM GAGE STATISTICS
SUM OF 3-DAY AVERAGE CFS

	Mean	Standard Deviation	1% 3-Day Volume (g = -.68)**
Aptos Creek*	2.477	.454	2012
Bodfish Creek*	2.219	.533	1547
Branciforte Creek*	3.001	.340	4170
Calabazas Creek* SCVWD #31	2.119	.418	759
Cedar Creek	2.252	.706	3448
Corralitos Creek*	2.573	.455	2520
Coyote Creek* near Gilroy	3.380	.499	19439
Coyote Creek at Madrone	3.405	.674	42889
Dry Creek	1.966	.682	1614
Fisher Creek* SCVWD #10	2.176	.732	3228
Hale Creek SCVWD #33	1.541	.814	1055
Los Animas Creek SCVWD #56	2.162	.701	2745
Los Gatos Creek*	3.285	.389	9848
Matadero Creek*	2.051	.476	827
Pacheco Creek	3.034	.890	45153
Packwood Creek SCVWD #57	1.469	.700	554
Permanente Creek* SCVWD #34	1.991	.718	1988
Redwood Creek*	1.821	.442	422

---

*The final 20 stations used for the peak regression analysis.

**For skew g = -.68 K = 1.821

TABLE 5 (Continued)

REGIONAL REGRESSION ANALYSIS
STREAM GAGE STATISTICS
SUM OF 3-DAY AVERAGE CFS

	Mean	Standard Deviation	1% 3-Day Volume (g = -.68)**
Ross Creek* SCVWD #21	1.784	.499	493
San Francisquito*	2.910	.581	9290
San Lorenzo Creek* at Hayward	2.809	.541	6225
San Lorenzo River*	3.729	.388	27261
San Tomas Creek SCVWD #29	1.910	.682	1419
Saratoga Creek* at Saratoga	2.555	.344	1518
Soquel Creek*	3.224	.434	10335
Upper Permanente SCVWD #32	1.794	.488	481
Uvas Creek* at Morgan Hill	3.272	.384	9359
Uvas Creek* above Uvas Reservoir	3.037	.457	7399
West Branch Soquel Creek*	2.774	.430	3606
Wildcat Creek SCVWD #30	2.148	.414	798
Zayante Creek	2.689	.437	3053

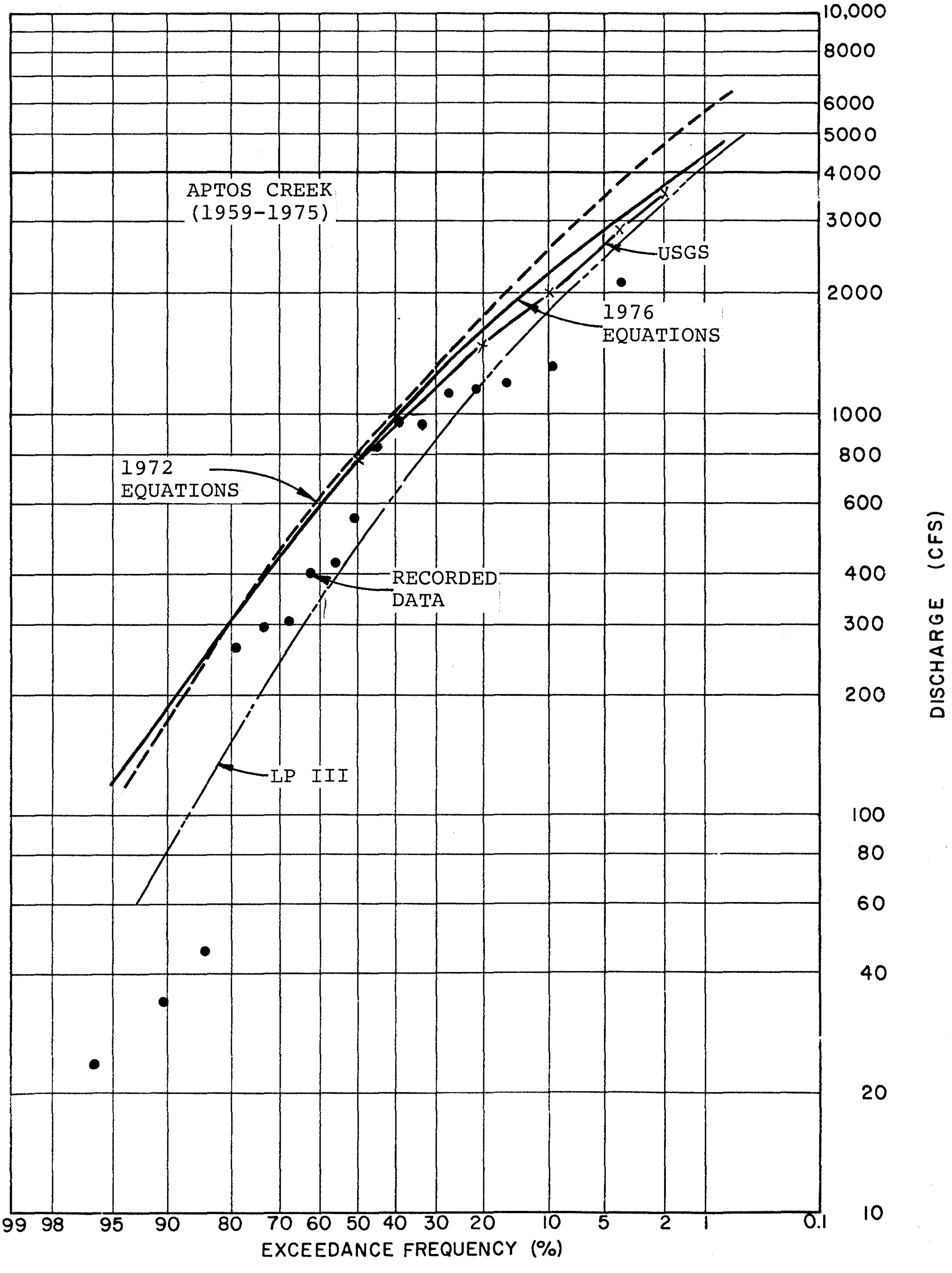
---

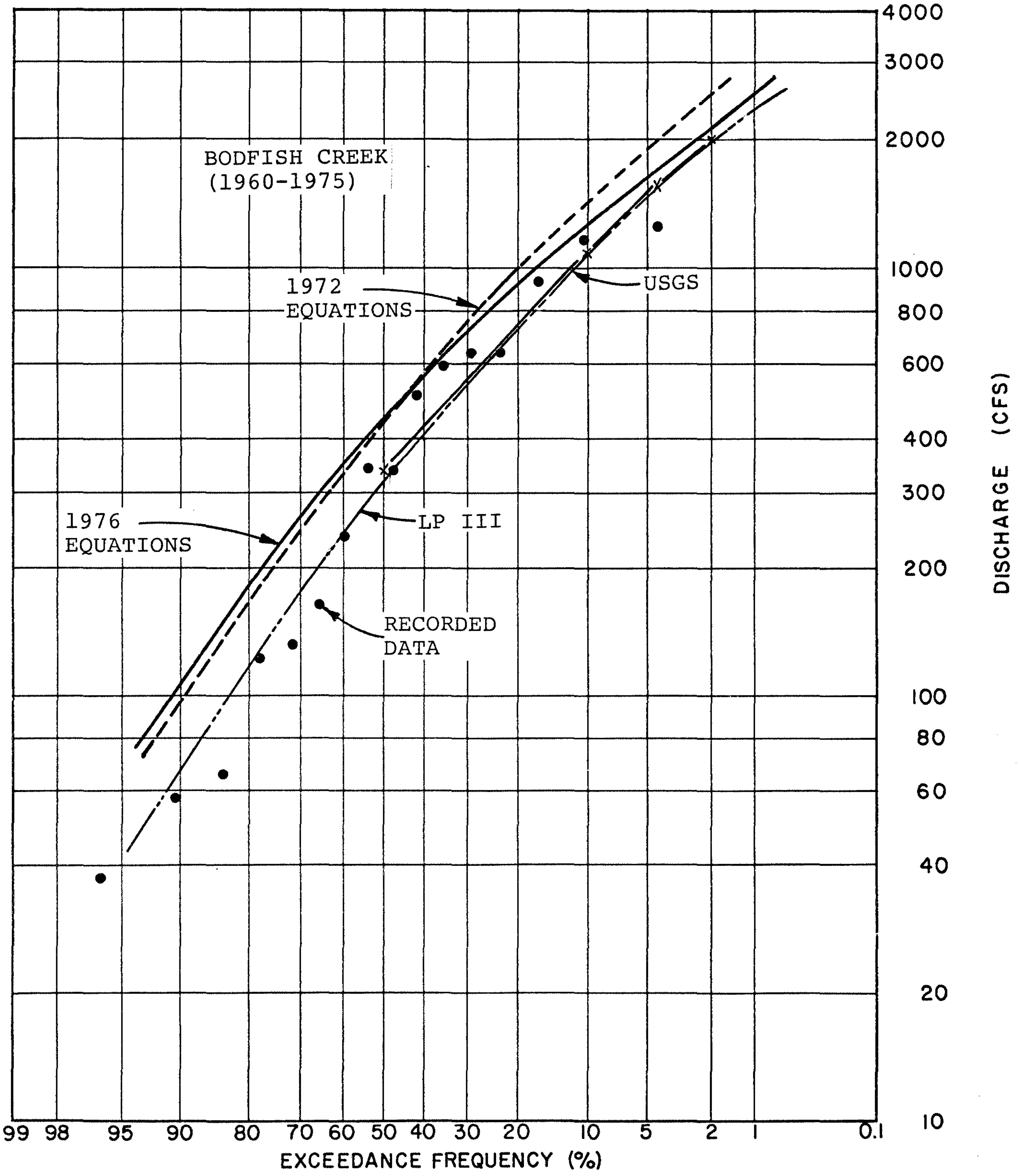
*The final 20 stations used for the peak regression analysis.

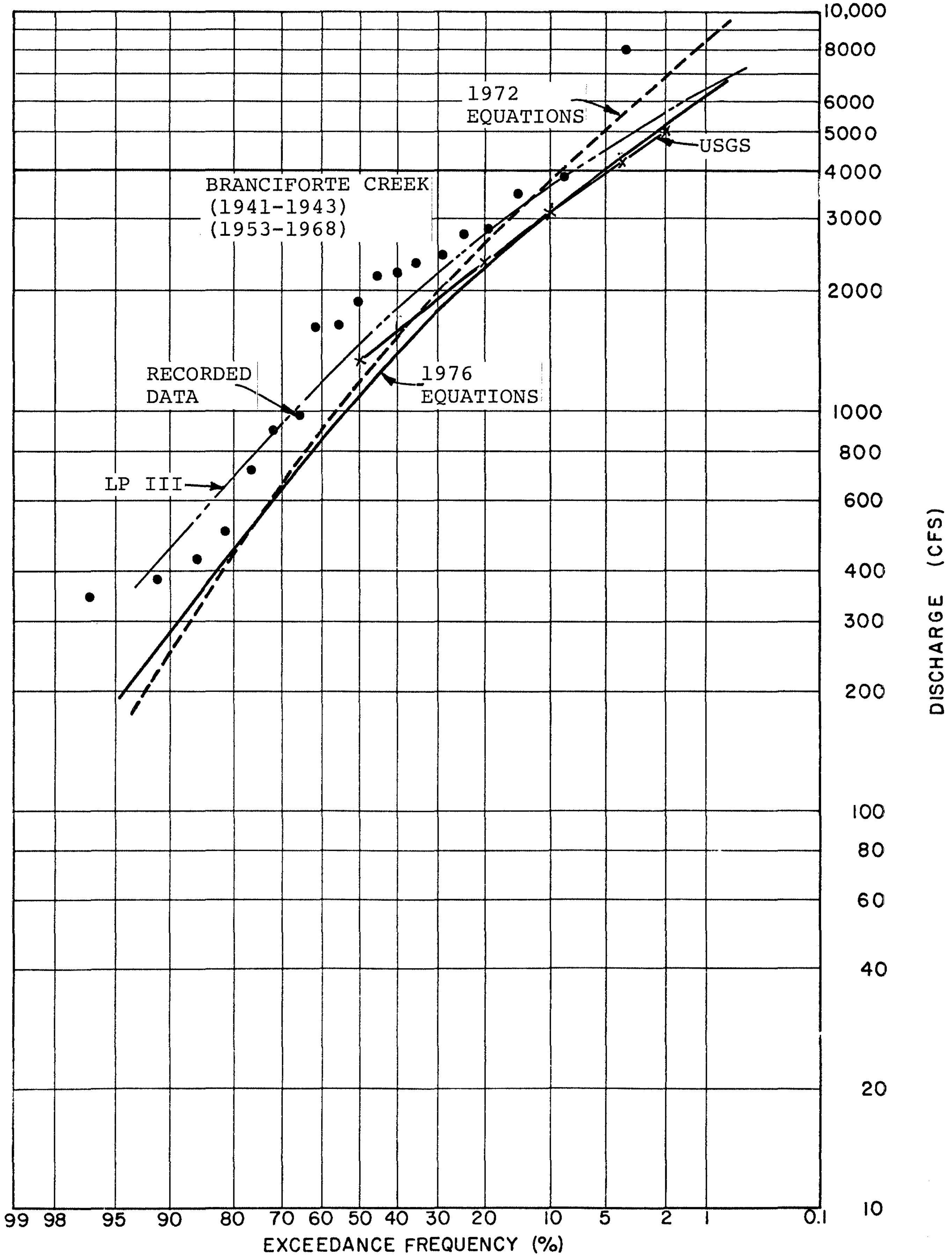
**For skew g = -.68 K = 1.821

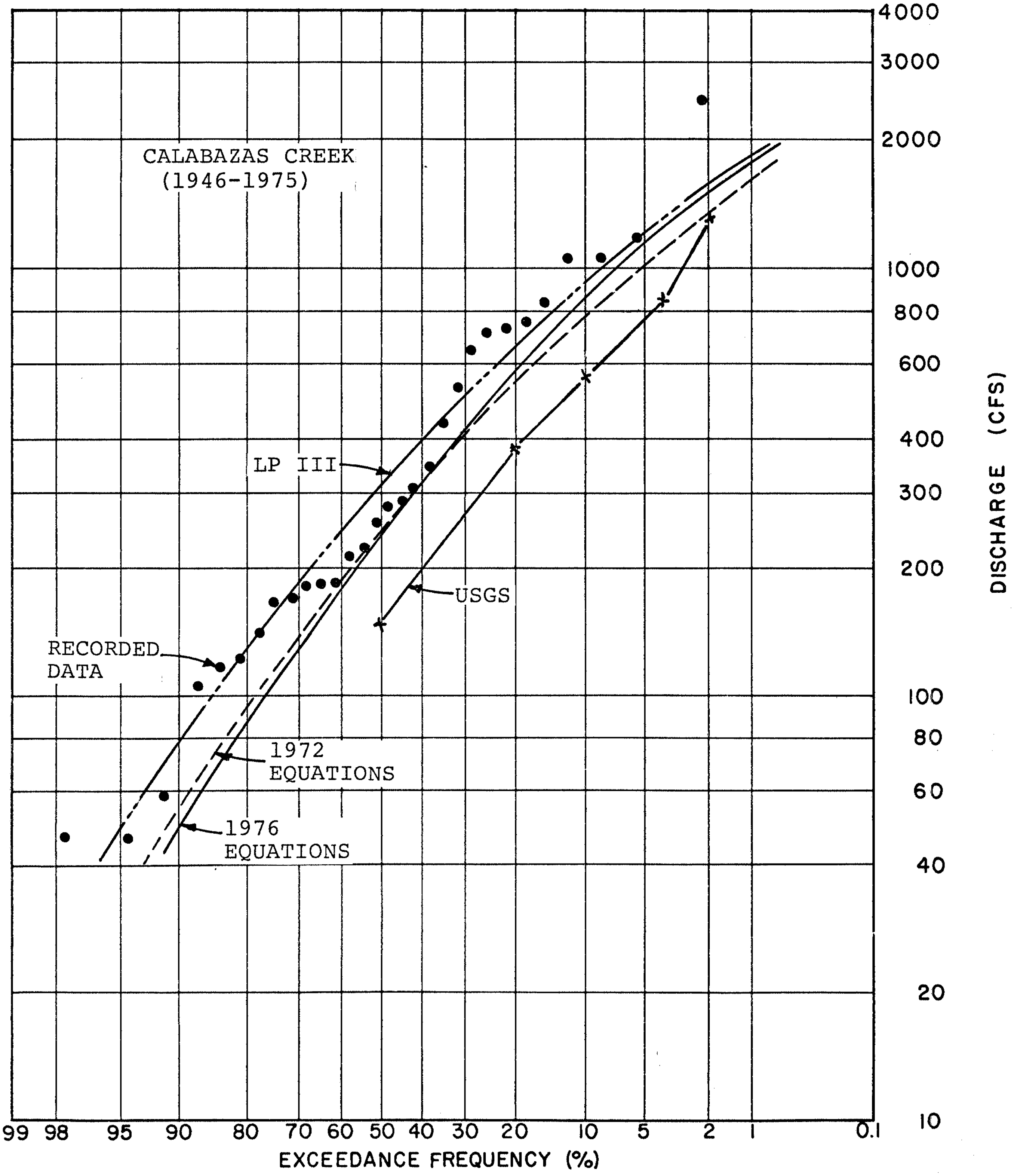
APPENDIX E

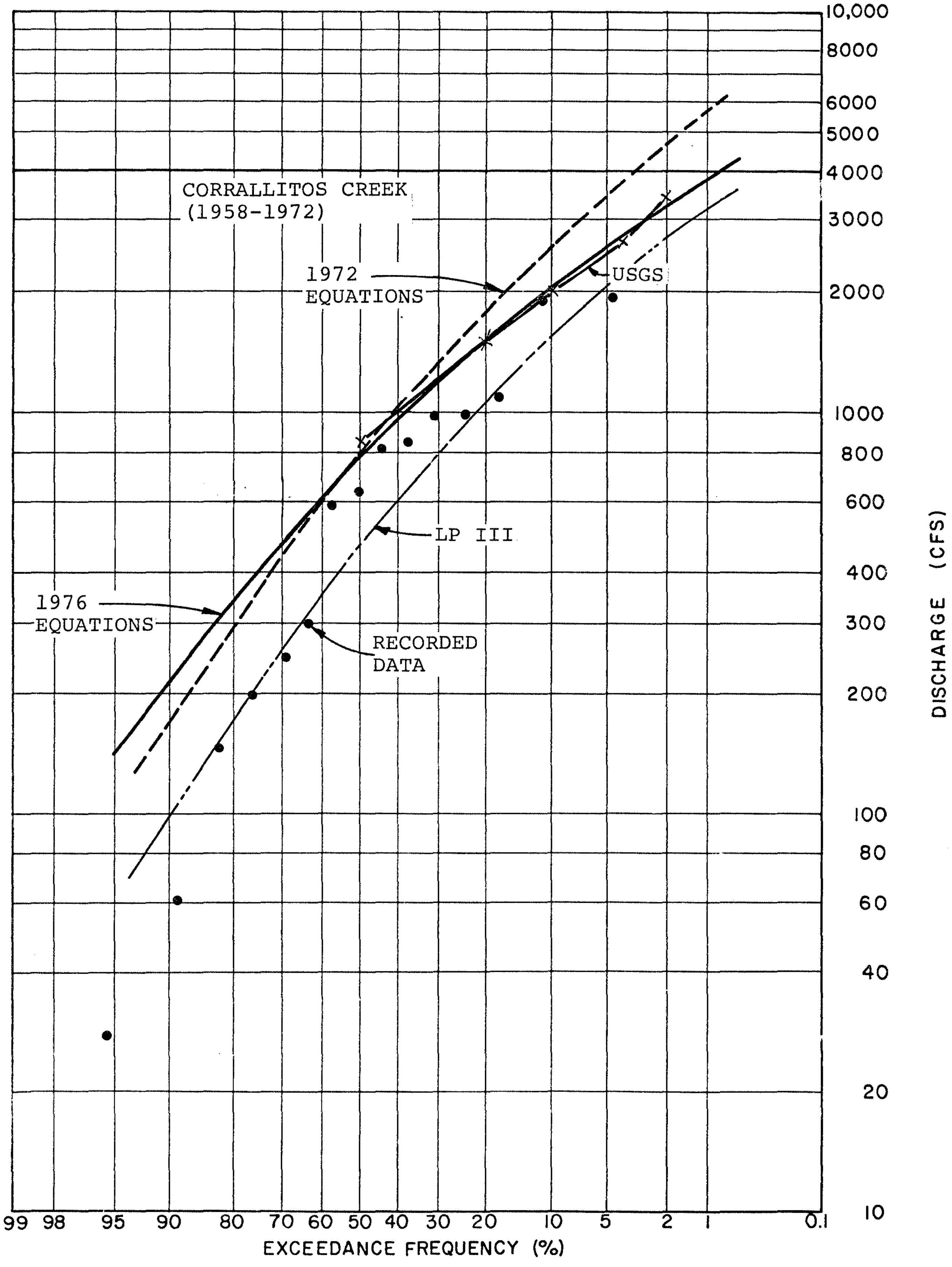
Frequency Curves

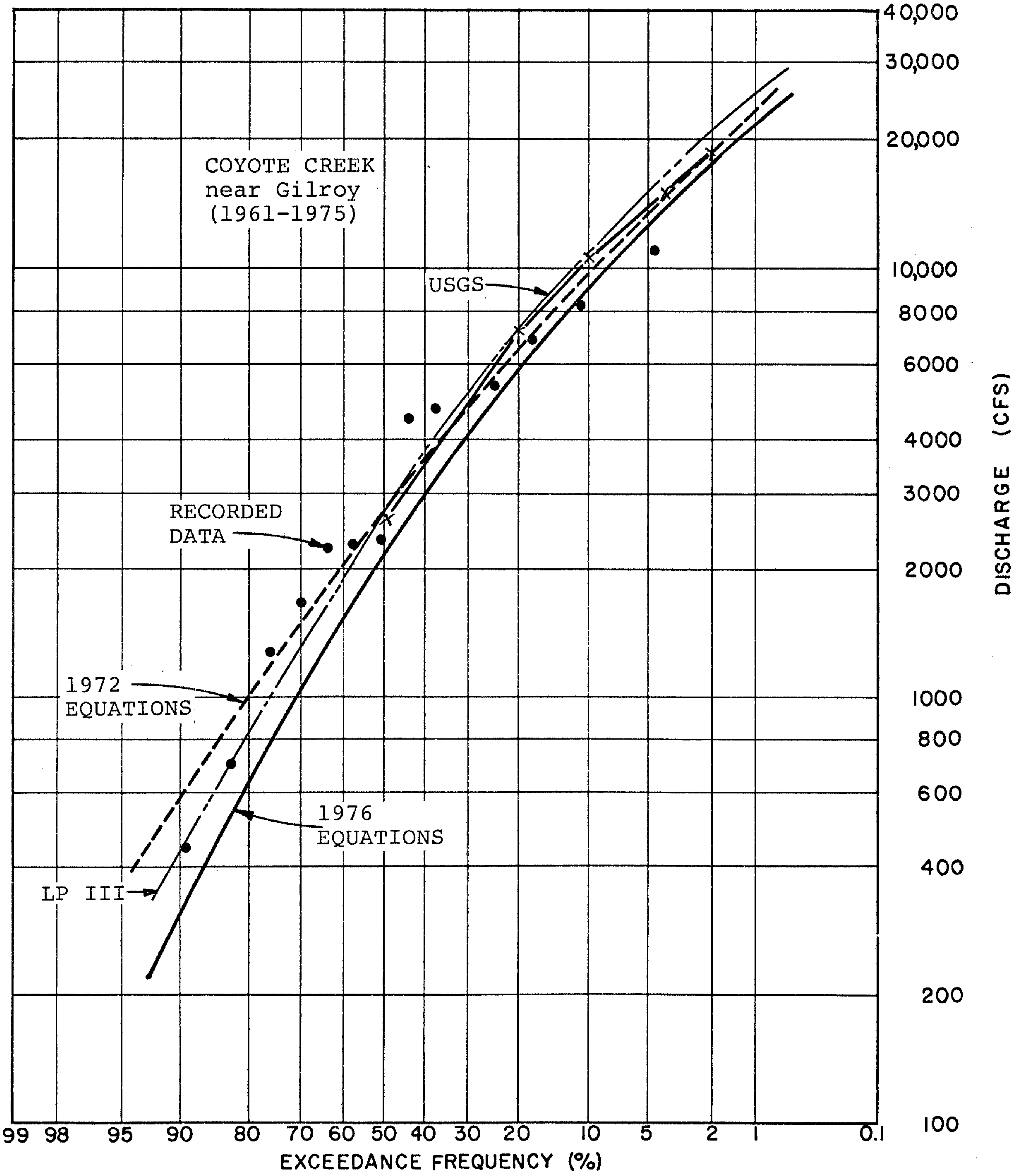

APTOS CREEK
(1959-1975)
USGS
1976
EQUATIONS
1972
EQUATIONS
RECORDED
DATA
LP III
10,000
8000
6000
5000
4000
3000
2000
1000
800
600
400
300
200
100
80
60
40
20
10
DISCHARGE (CFS)
99
98
95
90
80
70
60
50
40
30
20
10
5
2
1
0.1
EXCEEDANCE FREQUENCY (%)

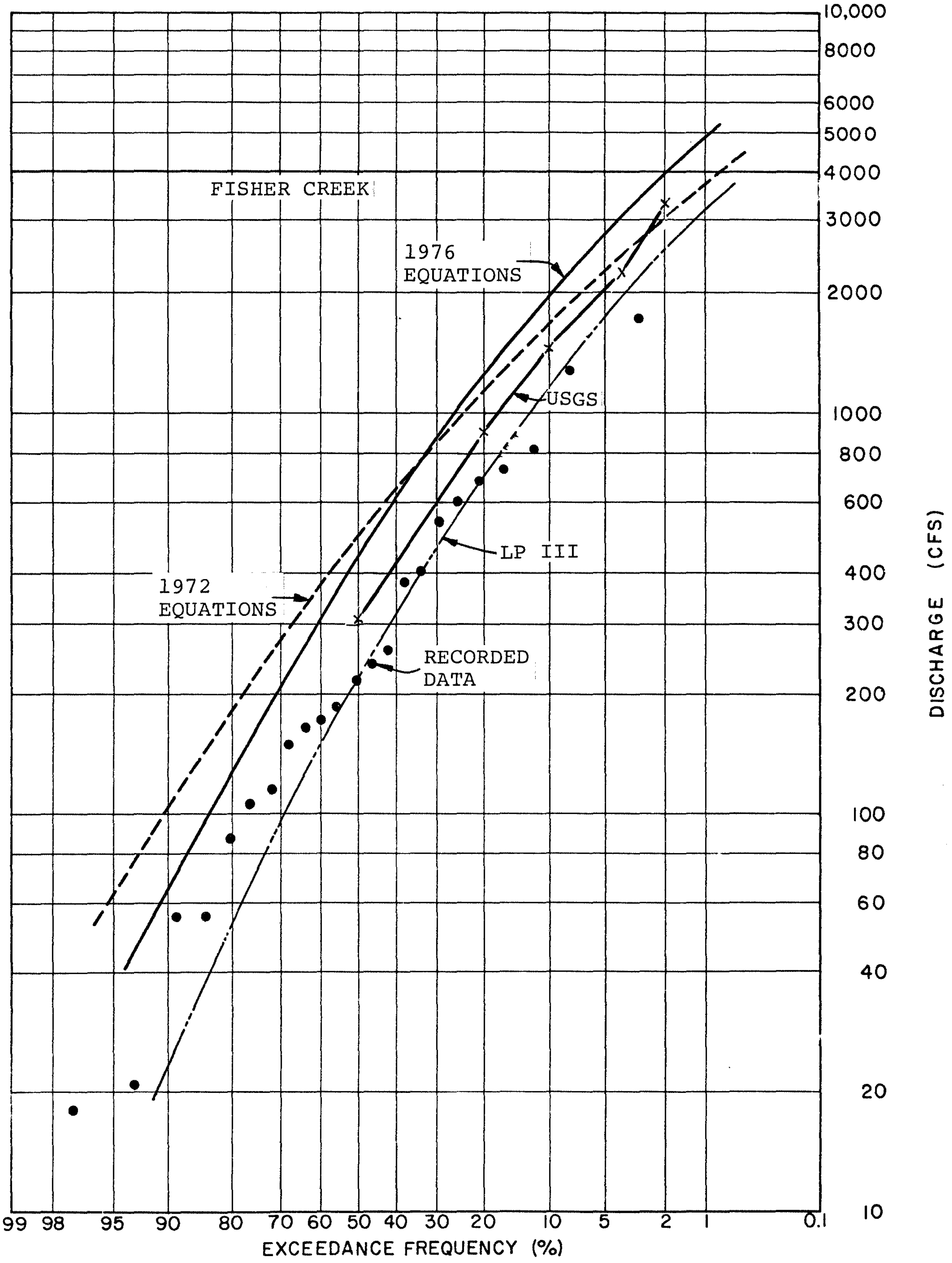

BODFISH CREEK
(1960-1975)
1972
EQUATIONS
USGS
1976
EQUATIONS
LP III
RECORDED
DATA
4000
3000
2000
1000
800
600
400
300
200
100
80
60
40
20
10
DISCHARGE (CFS)
99 98 95 90 80 70 60 50 40 30 20 10 5 2 1 0.1
EXCEEDANCE FREQUENCY (%)

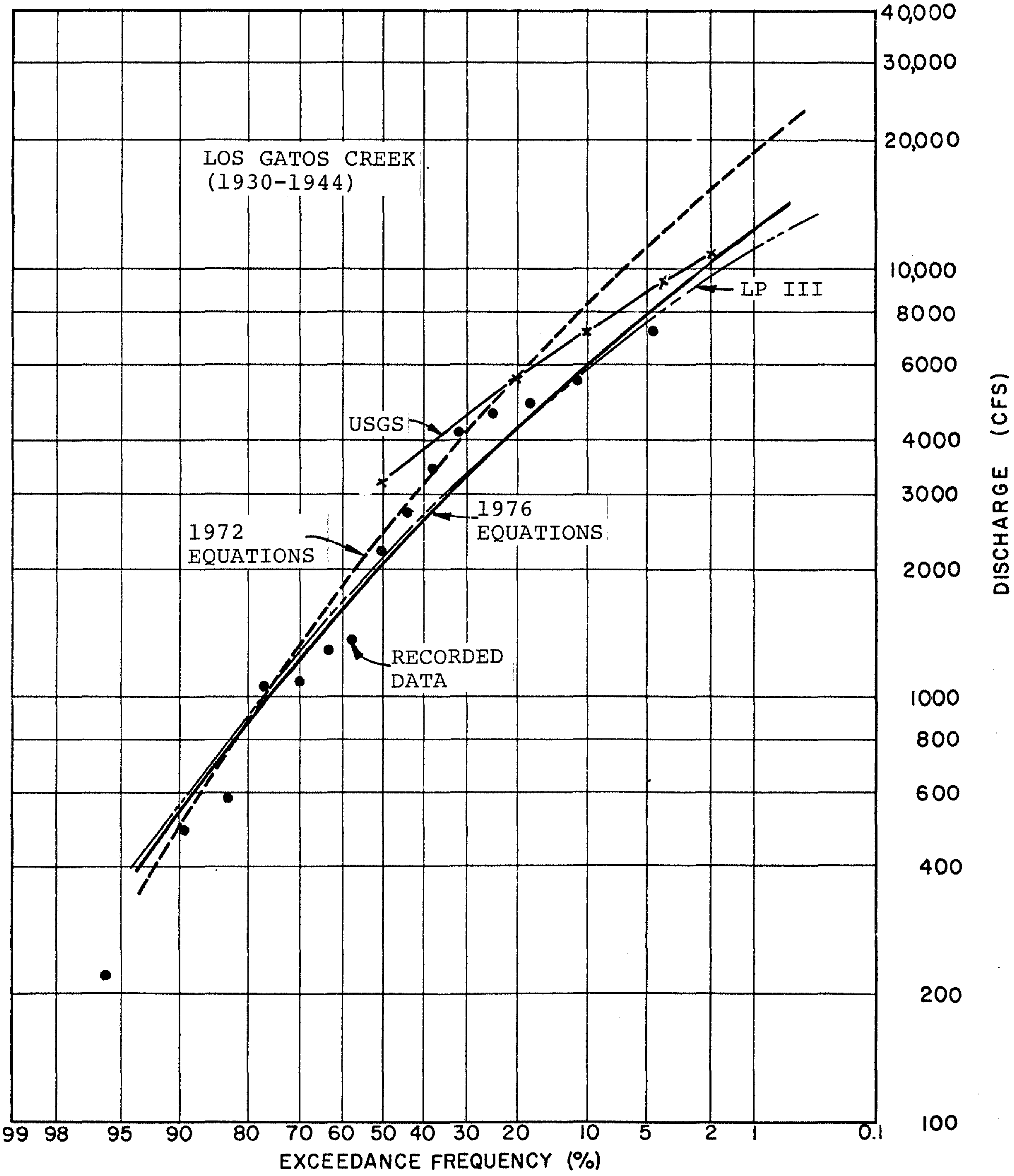

BRANCIFORTE CREEK
(1941-1943)
(1953-1968)
1972 EQUATIONS
USGS
RECORDED DATA
1976 EQUATIONS
LP III
10,000
8000
6000
5000
4000
3000
2000
1000
800
600
400
300
200
100
80
60
40
20
10
DISCHARGE (CFS)
99
98
95
90
80
70
60
50
40
30
20
10
5
2
1
0.1
EXCEEDANCE FREQUENCY (%)

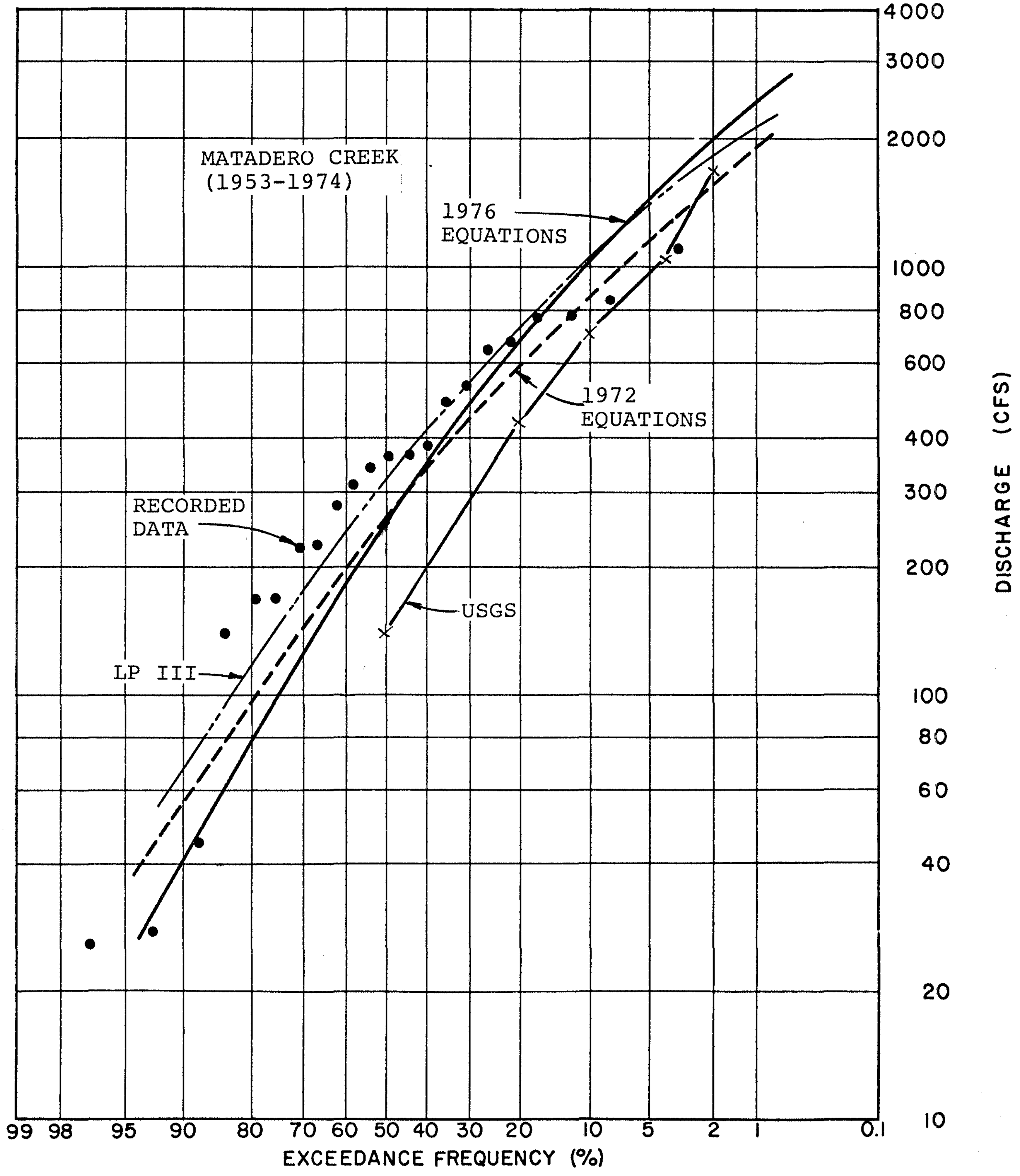

CALABAZAS CREEK
(1946-1975)
LP III
USGS
RECORDED DATA
1972 EQUATIONS
1976 EQUATIONS
99 98 95 90 80 70 60 50 40 30 20 10 5 2 1 0.1
EXCEEDANCE FREQUENCY (%)
4000 3000 2000 1000 800 600 400 300 200 100 80 60 40 20 10
DISCHARGE (CFS)

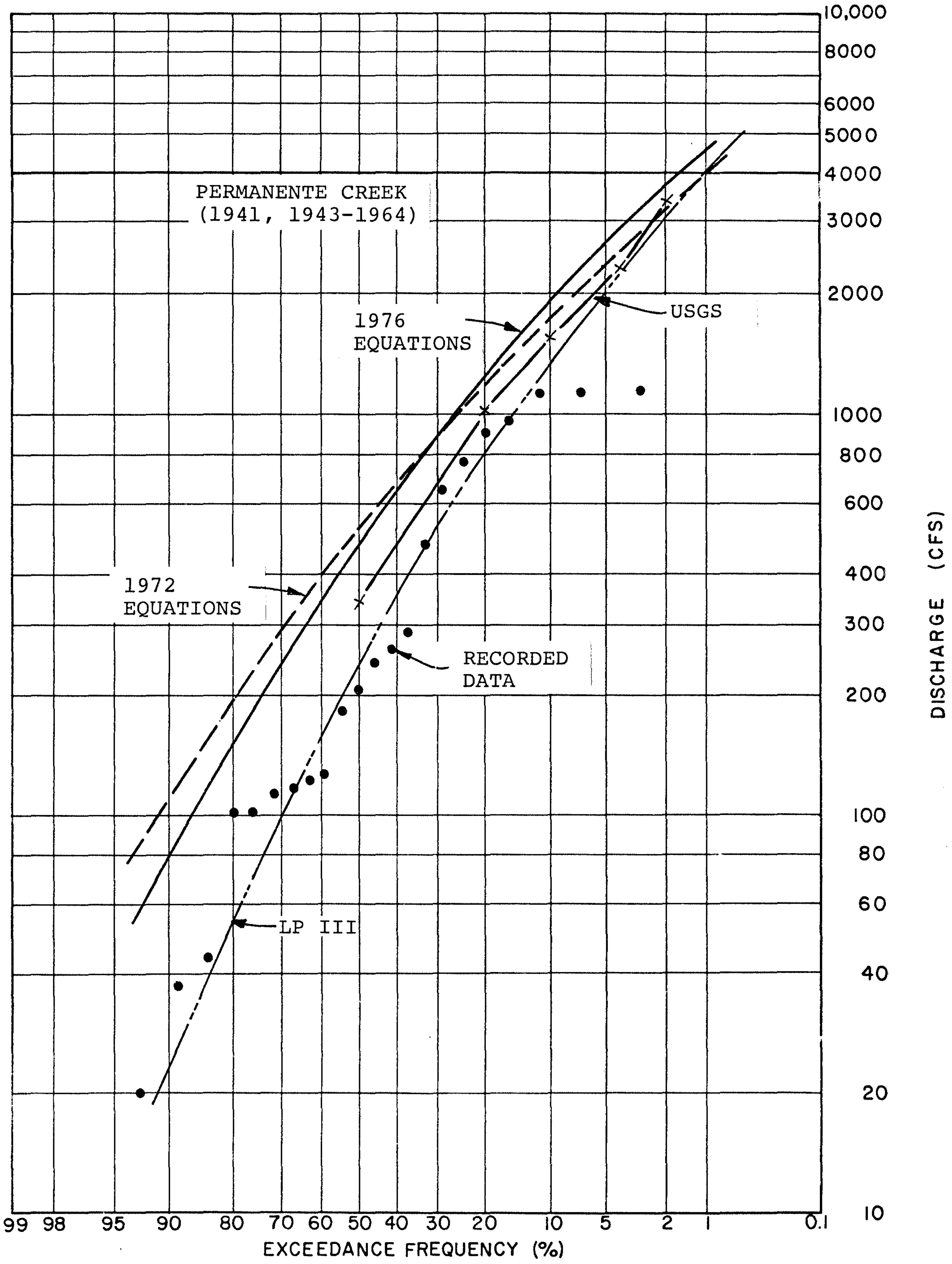

CORRALLITOS CREEK
(1958-1972)
1972
EQUATIONS
USGS
LP III
1976
EQUATIONS
RECORDED
DATA
10,000
8000
6000
5000
4000
3000
2000
1000
800
600
400
300
200
100
80
60
40
20
10
DISCHARGE (CFS)
99 98 95 90 80 70 60 50 40 30 20 10 5 2 1 0.1
EXCEEDANCE FREQUENCY (%)

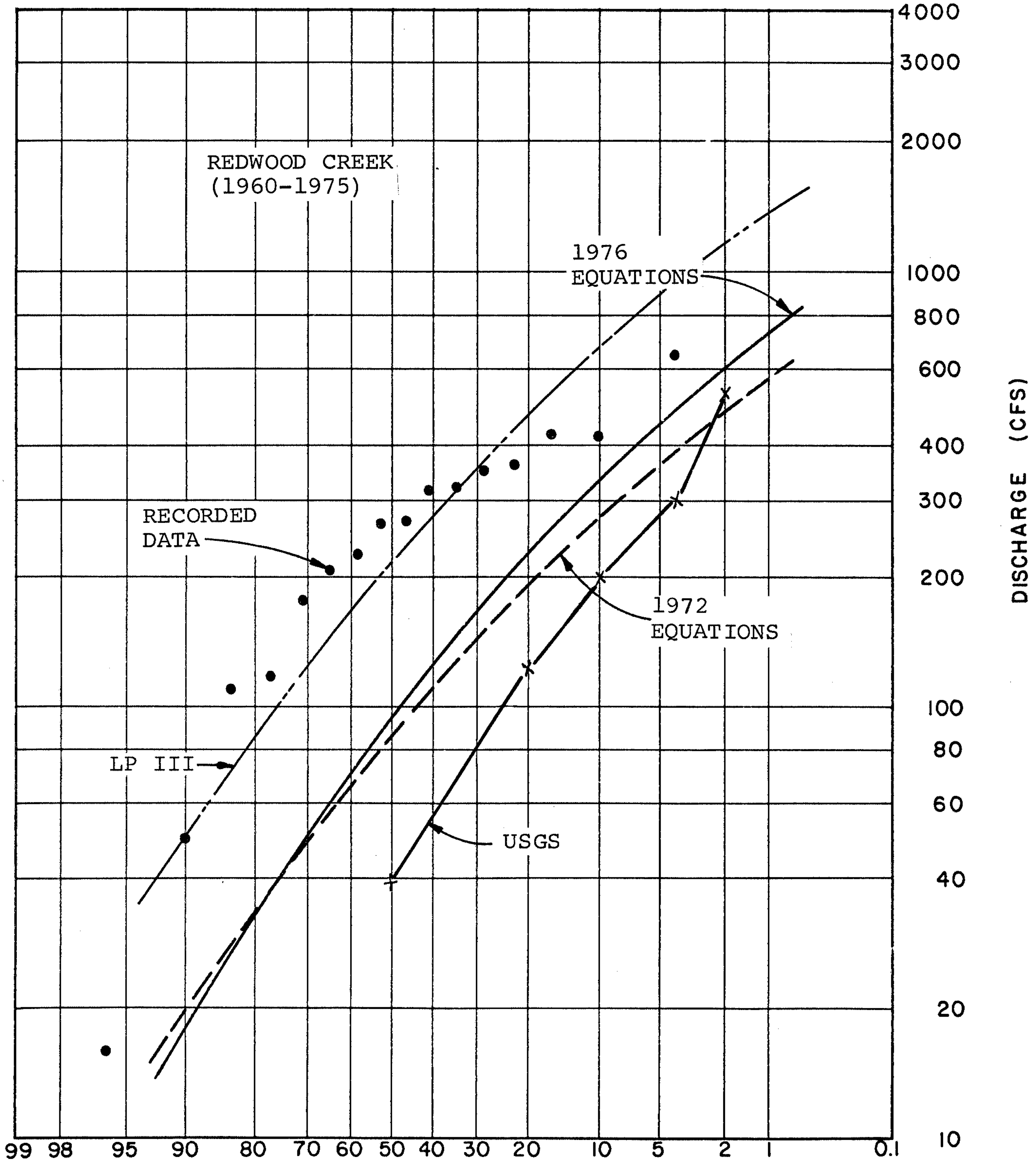

COYOTE CREEK
near Gilroy
(1961-1975)
USGS
RECORDED DATA
1972 EQUATIONS
1976 EQUATIONS
LP III
40,000
30,000
20,000
10,000
8000
6000
4000
3000
2000
1000
800
600
400
200
100
DISCHARGE (CFS)
99
98
95
90
80
70
60
50
40
30
20
10
5
2
1
0.1
EXCEEDANCE FREQUENCY (%)

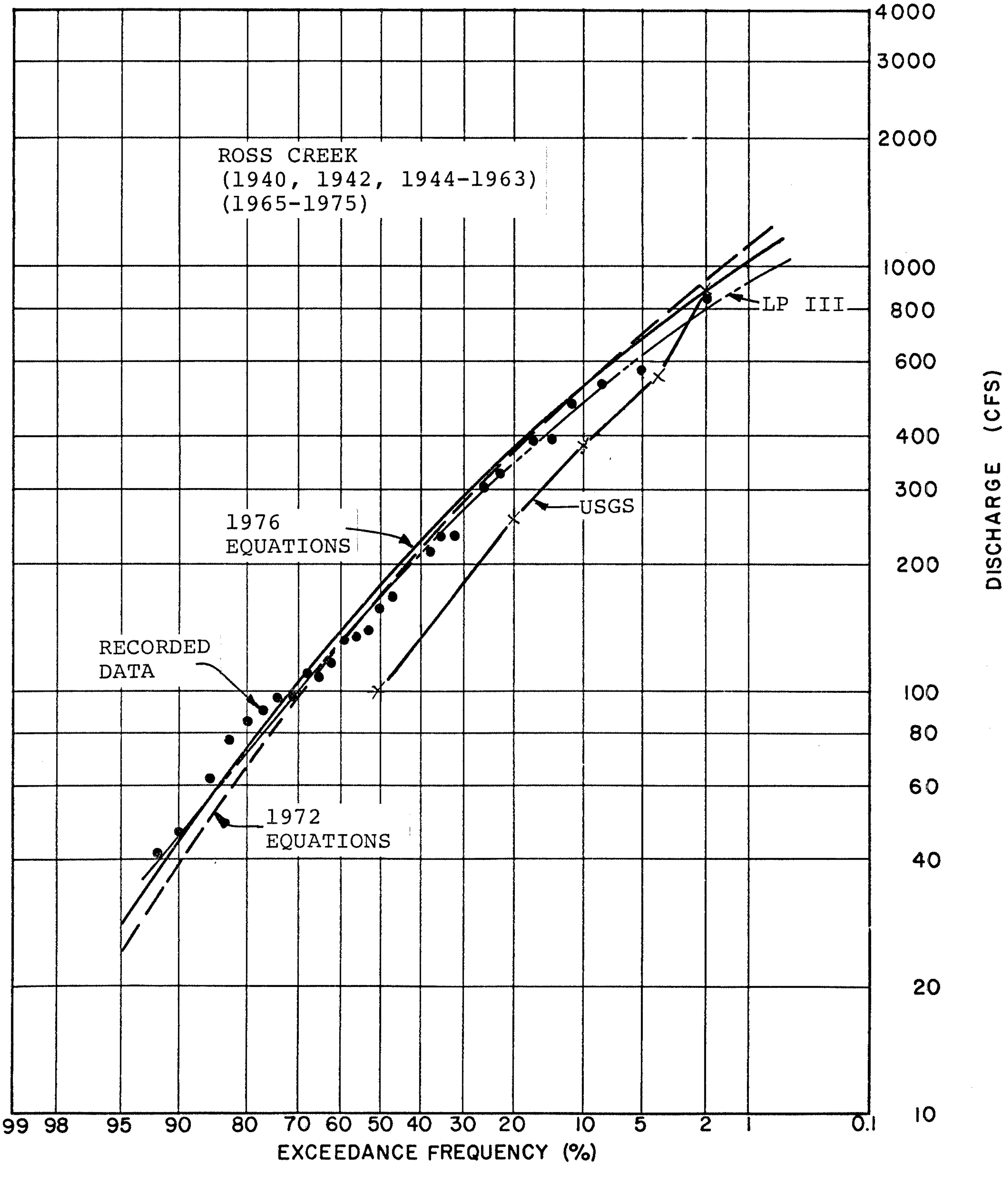

FISHER CREEK
1976 EQUATIONS
USGS
LP III
1972 EQUATIONS
RECORDED DATA
10,000
8000
6000
5000
4000
3000
2000
1000
800
600
400
300
200
100
80
60
40
20
10
DISCHARGE (CFS)
99 98 95 90 80 70 60 50 40 30 20 10 5 2 1 0.1
EXCEEDANCE FREQUENCY (%)

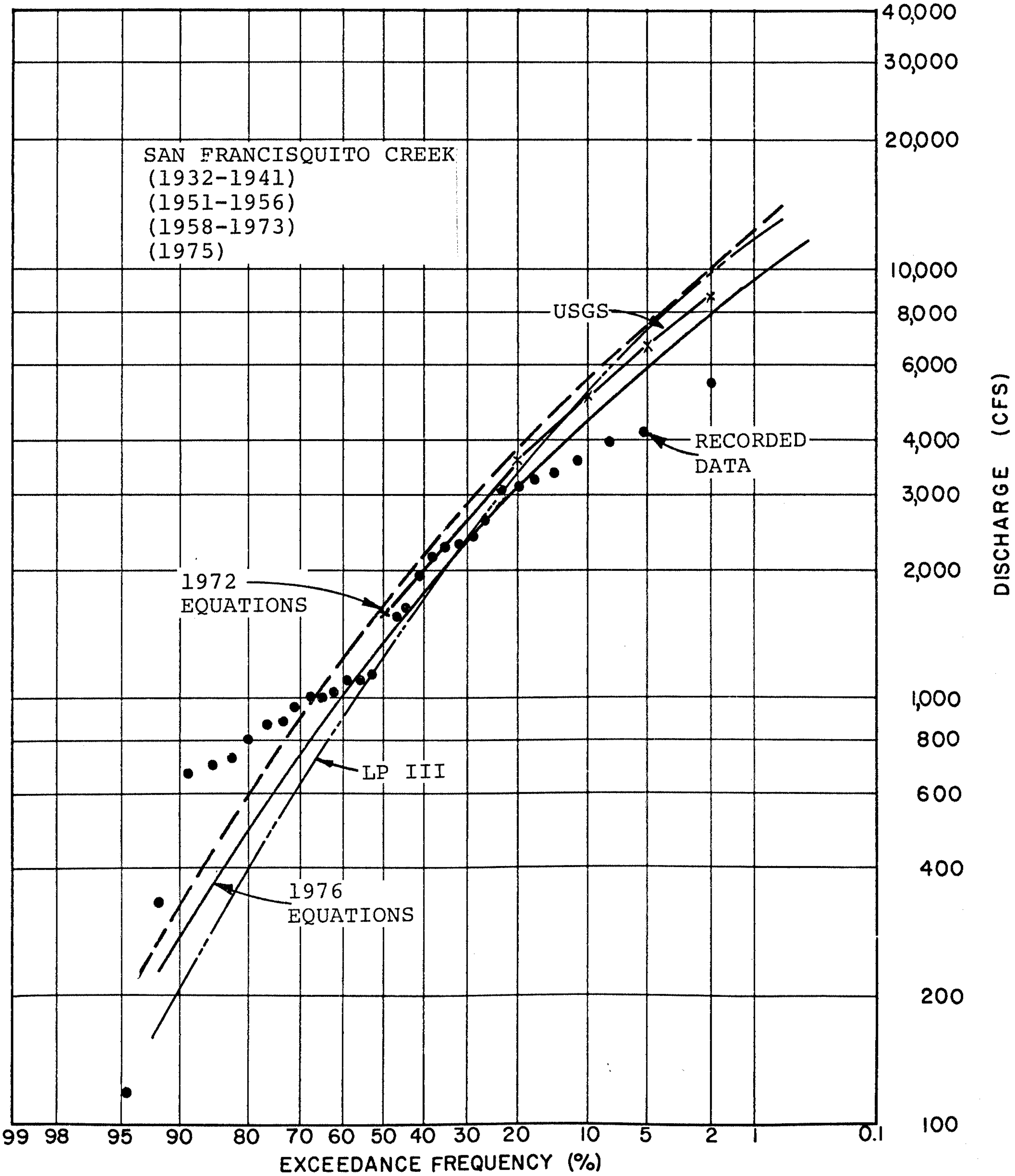

LOS GATOS CREEK
(1930-1944)
USGS
LP III
1972
EQUATIONS
1976
EQUATIONS
RECORDED
DATA
40,000
30,000
20,000
10,000
8000
6000
4000
3000
2000
1000
800
600
400
200
100
DISCHARGE (CFS)
99 98 95 90 80 70 60 50 40 30 20 10 5 2 1 0.1
EXCEEDANCE FREQUENCY (%)

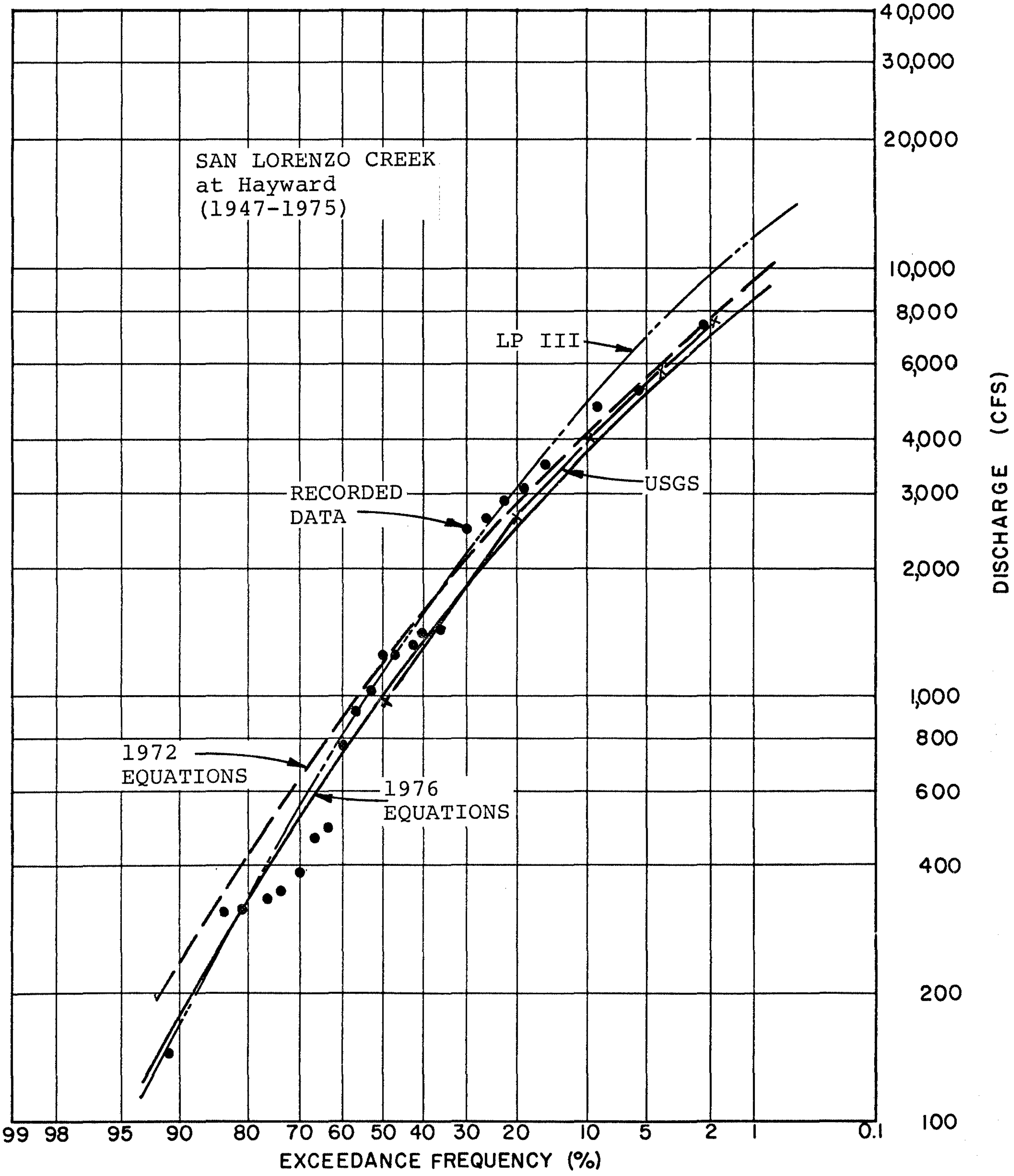

MATADERO CREEK
(1953-1974)
1976
EQUATIONS
1972
EQUATIONS
RECORDED
DATA
USGS
LP III
4000
3000
2000
1000
800
600
400
300
200
100
80
60
40
20
10
DISCHARGE (CFS)
99 98 95 90 80 70 60 50 40 30 20 10 5 2 1 0.1
EXCEEDANCE FREQUENCY (%)

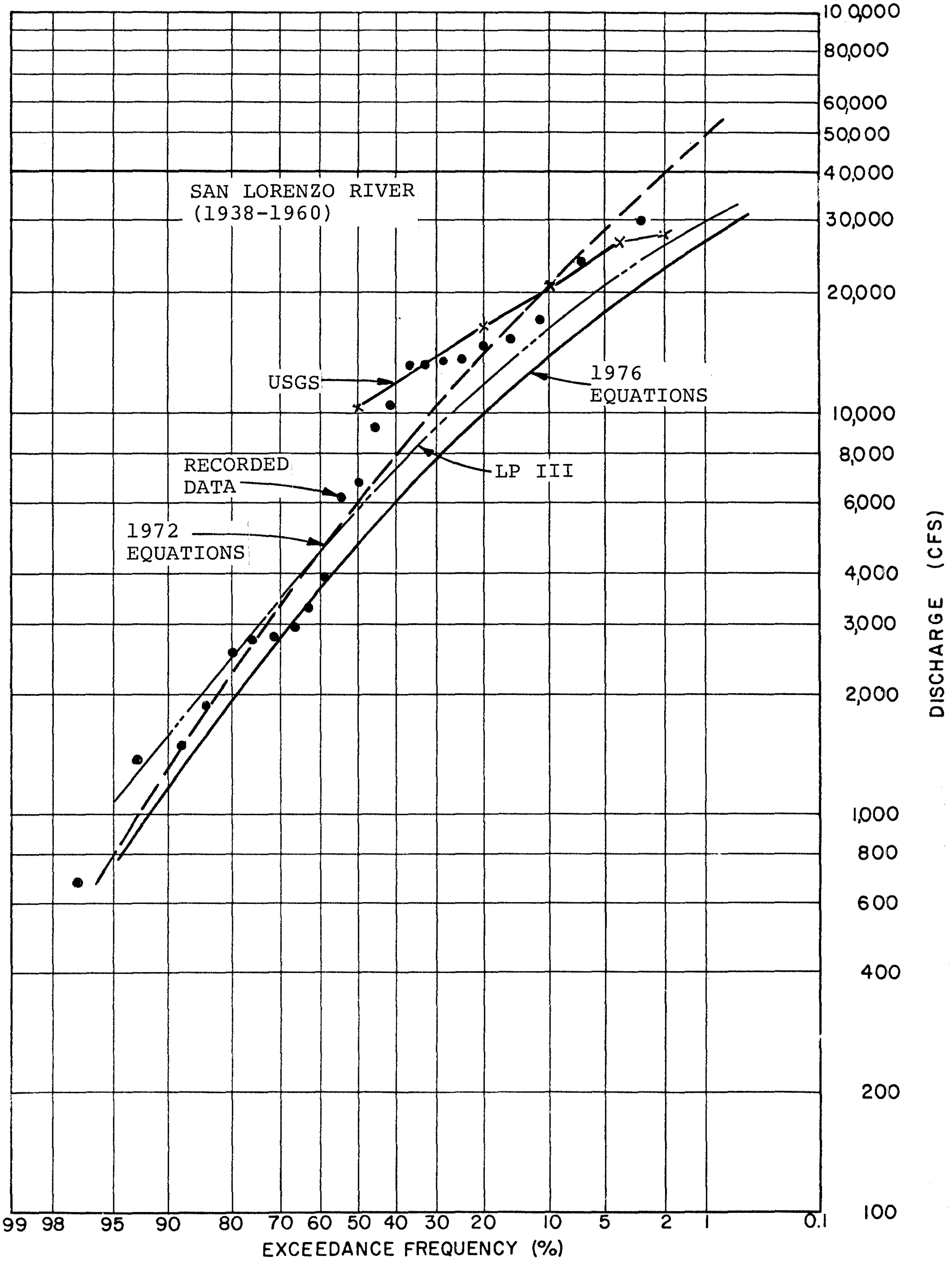

PERMANENTE CREEK
(1941, 1943-1964)
1976
EQUATIONS
USGS
1972
EQUATIONS
RECORDED
DATA
LP III
10,000
8000
6000
5000
4000
3000
2000
1000
800
600
400
300
200
100
80
60
40
20
10
DISCHARGE (CFS)
99
98
95
90
80
70
60
50
40
30
20
10
5
2
1
0.1
EXCEEDANCE FREQUENCY (%)

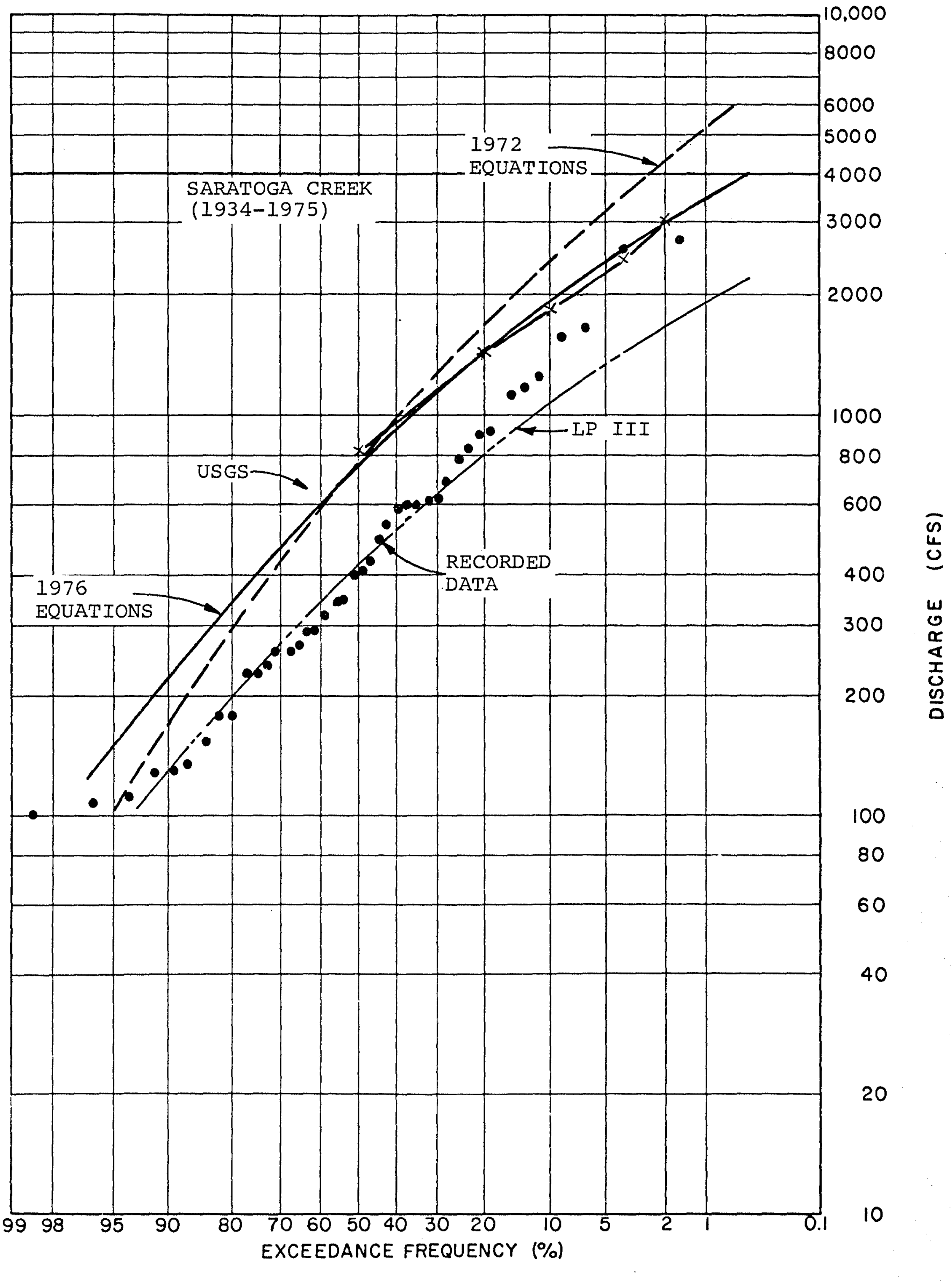

REDWOOD CREEK
(1960-1975)
1976
EQUATIONS
RECORDED
DATA
1972
EQUATIONS
LP III
USGS
4000
3000
2000
1000
800
600
400
300
200
100
80
60
40
20
10
DISCHARGE (CFS)
99 98 95 90 80 70 60 50 40 30 20 10 5 2 1 0.1
EXCEEDANCE FREQUENCY (%)

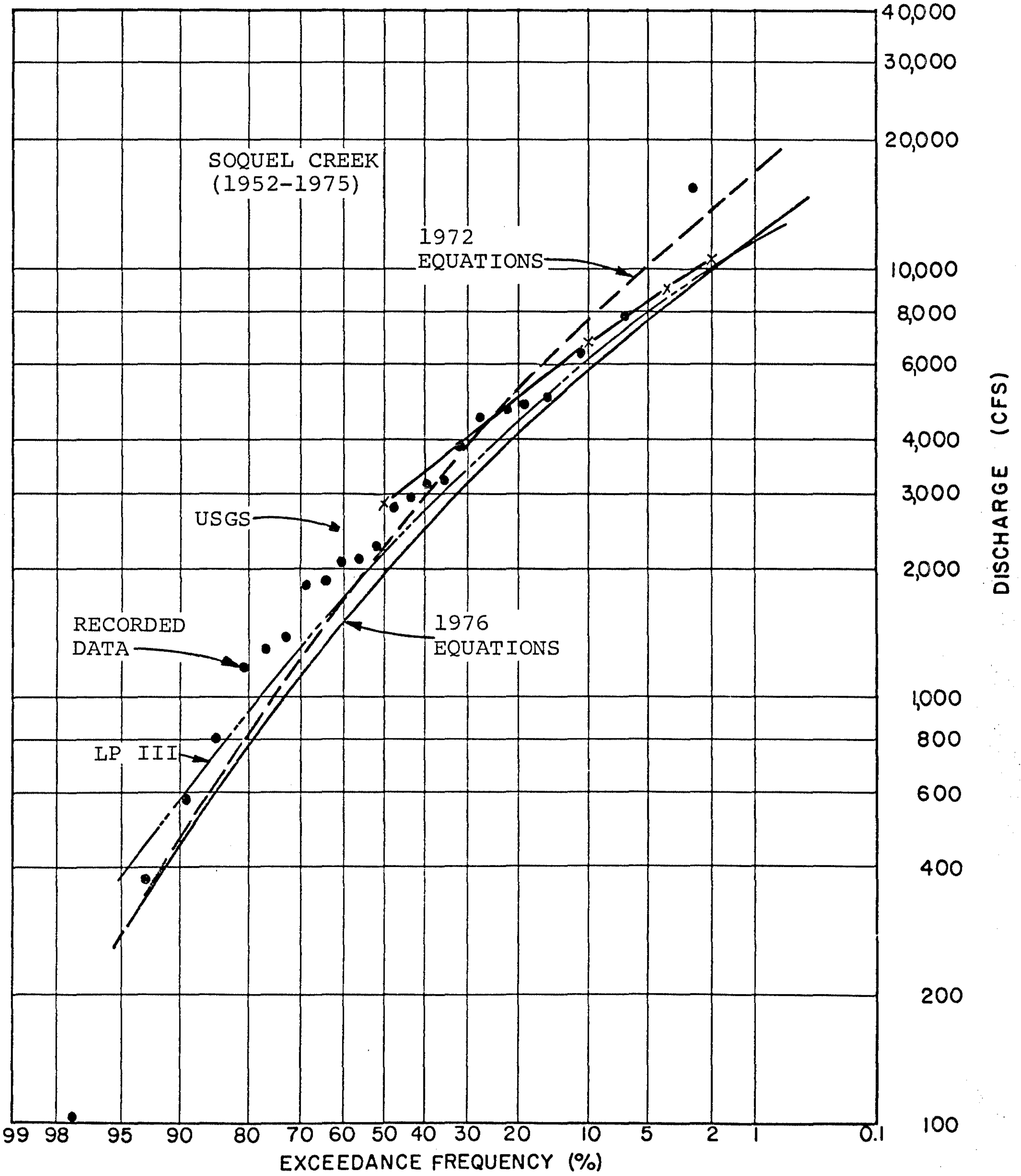

ROSS CREEK
(1940, 1942, 1944-1963)
(1965-1975)
1976 EQUATIONS
RECORDED DATA
1972 EQUATIONS
USGS
LP III
4000
3000
2000
1000
800
600
400
300
200
100
80
60
40
20
10
DISCHARGE (CFS)
99
98
95
90
80
70
60
50
40
30
20
10
5
2
1
0.1
EXCEEDANCE FREQUENCY (%)

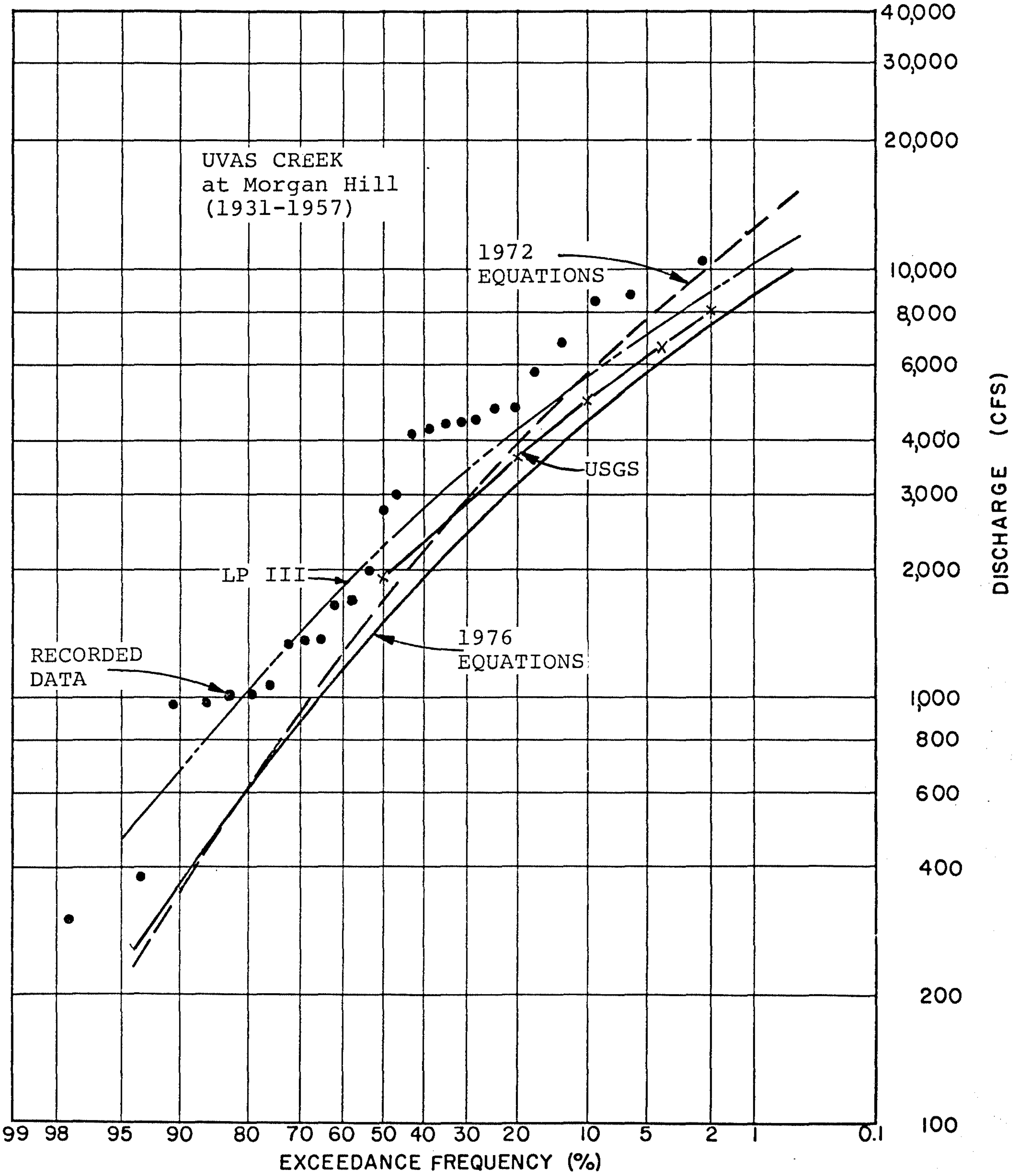

SAN FRANCISQUITO CREEK
(1932-1941)
(1951-1956)
(1958-1973)
(1975)
USGS
RECORDED DATA
1972 EQUATIONS
LP III
1976 EQUATIONS
40,000
30,000
20,000
10,000
8,000
6,000
4,000
3,000
2,000
1,000
800
600
400
200
100
DISCHARGE (CFS)
99 98 95 90 80 70 60 50 40 30 20 10 5 2 1 0.1
EXCEEDANCE FREQUENCY (%)

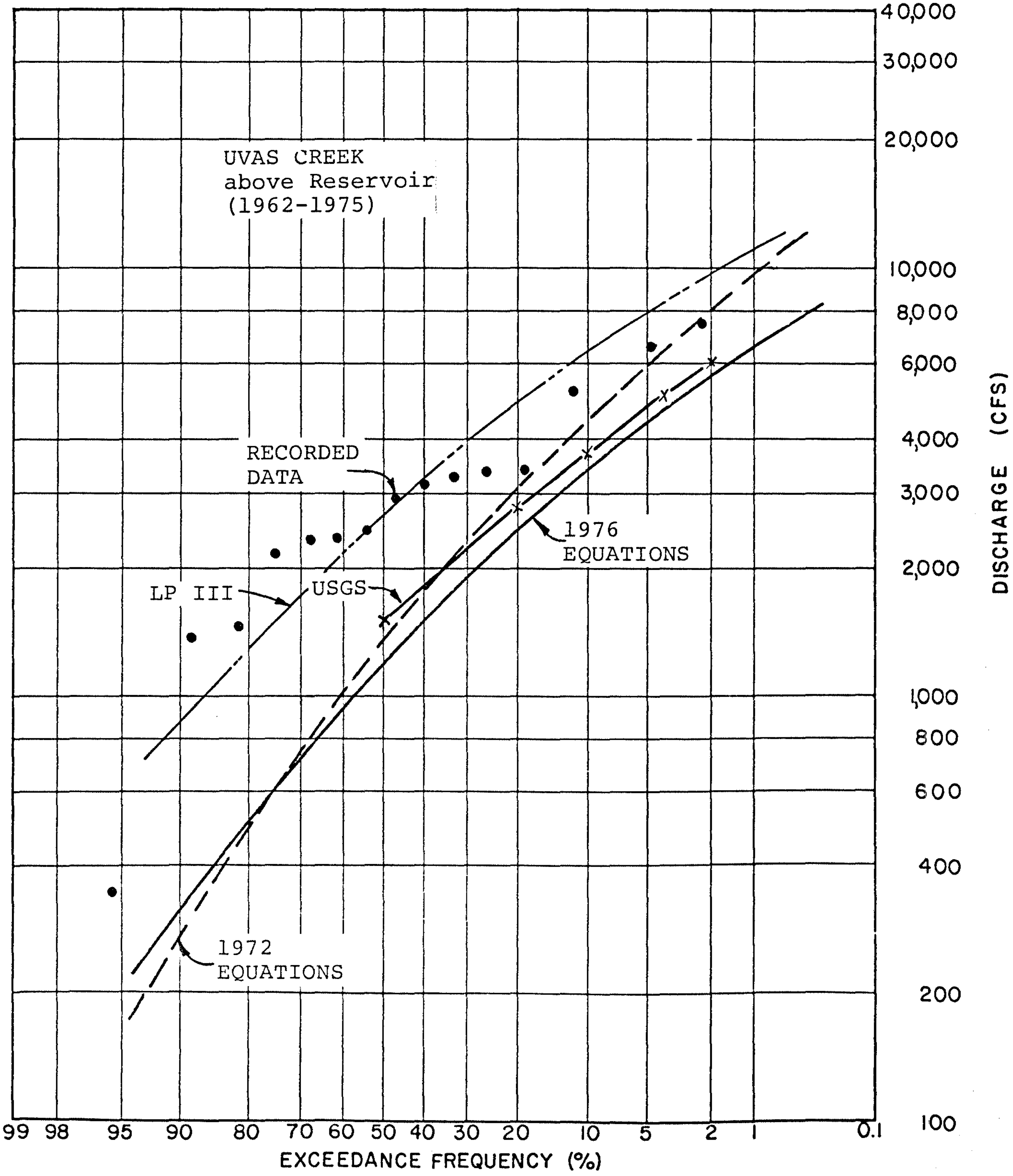

SAN LORENZO CREEK
at Hayward
(1947-1975)
LP III
USGS
RECORDED DATA
1972 EQUATIONS
1976 EQUATIONS
40,000
30,000
20,000
10,000
8,000
6,000
4,000
3,000
2,000
1,000
800
600
400
200
100
DISCHARGE (CFS)
99
98
95
90
80
70
60
50
40
30
20
10
5
2
1
0.1
EXCEEDANCE FREQUENCY (%)


SAN LORENZO RIVER
(1938-1960)
USGS
RECORDED DATA
1972 EQUATIONS
1976 EQUATIONS
LP III
10 0,000
80,000
60,000
50,000
40,000
30,000
20,000
10,000
8,000
6,000
4,000
3,000
2,000
1,000
800
600
400
200
100
DISCHARGE (CFS)
99
98
95
90
80
70
60
50
40
30
20
10
5
2
1
0.1
EXCEEDANCE FREQUENCY (%)

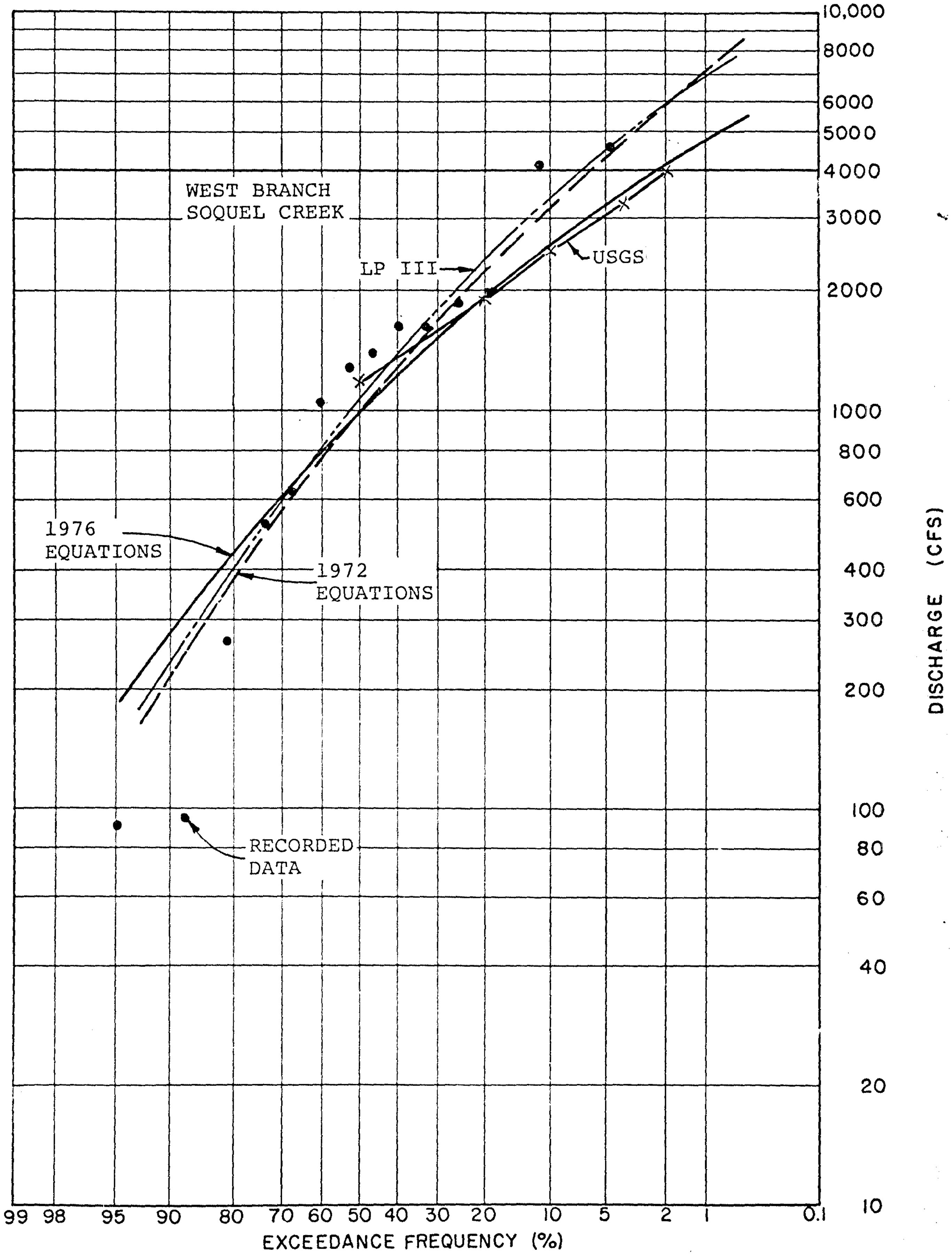

SARATOGA CREEK
(1934-1975)
1972
EQUATIONS
USGS
LP III
RECORDED
DATA
1976
EQUATIONS
10,000
8000
6000
5000
4000
3000
2000
1000
800
600
400
300
200
100
80
60
40
20
10
DISCHARGE (CFS)
99
98
95
90
80
70
60
50
40
30
20
10
5
2
1
0.1
EXCEEDANCE FREQUENCY (%)


SOQUEL CREEK
(1952-1975)
1972
EQUATIONS
USGS
RECORDED
DATA
1976
EQUATIONS
LP III
40,000
30,000
20,000
10,000
8,000
6,000
4,000
3,000
2,000
1,000
800
600
400
200
100
DISCHARGE (CFS)
99 98 95 90 80 70 60 50 40 30 20 10 5 2 1 0.1
EXCEEDANCE FREQUENCY (%)


UVAS CREEK
at Morgan Hill
(1931-1957)
1972 EQUATIONS
USGS
LP III
1976 EQUATIONS
RECORDED DATA
40,000
30,000
20,000
10,000
8,000
6,000
4,000
3,000
2,000
1,000
800
600
400
200
100
DISCHARGE (CFS)
99 98 95 90 80 70 60 50 40 30 20 10 5 2 1 0.1
EXCEEDANCE FREQUENCY (%)



UVAS CREEK
above Reservoir
(1962-1975)
RECORDED DATA
LP III
USGS
1976 EQUATIONS
1972 EQUATIONS
40,000
30,000
20,000
10,000
8,000
6,000
4,000
3,000
2,000
1,000
800
600
400
200
100
DISCHARGE (CFS)
99
98
95
90
80
70
60
50
40
30
20
10
5
2
1
0.1
EXCEEDANCE FREQUENCY (%)


WEST BRANCH
SOQUEL CREEK
LP III
USGS
1976
EQUATIONS
1972
EQUATIONS
RECORDED
DATA
10,000
8000
6000
5000
4000
3000
2000
1000
800
600
400
300
200
100
80
60
40
20
10
DISCHARGE (CFS)
99
98
95
90
80
70
60
50
40
30
20
10
5
2
1
0.1
EXCEEDANCE FREQUENCY (%)